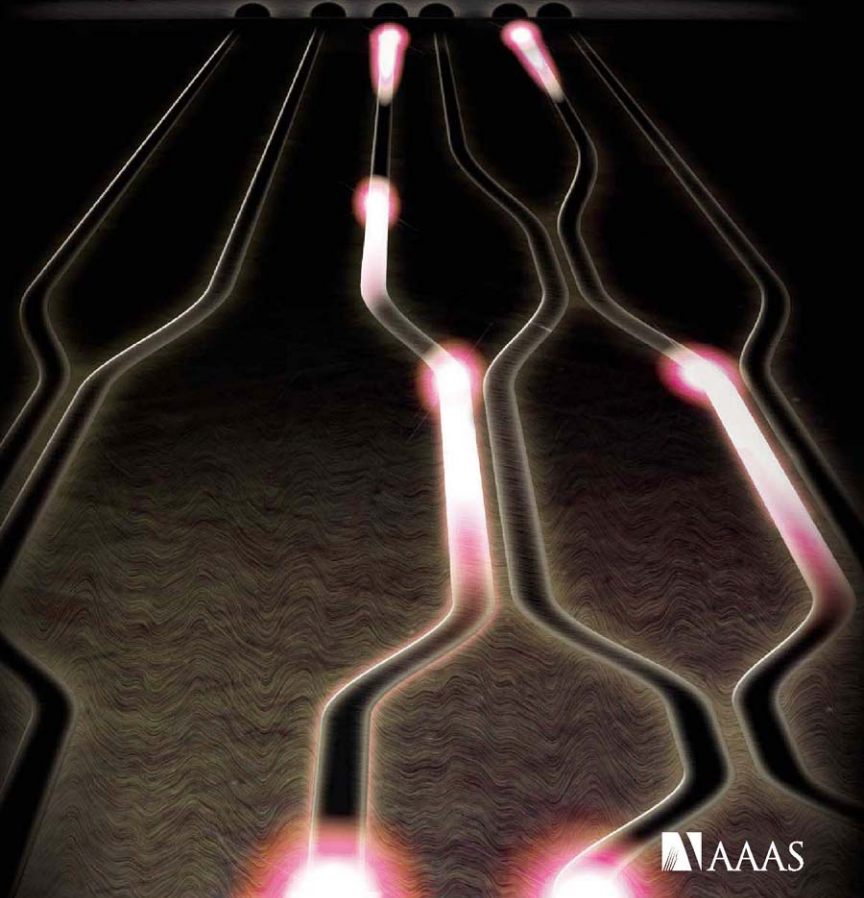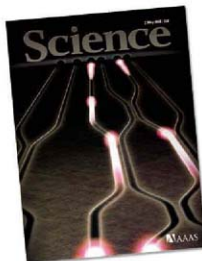# Science

COVER

Graphic representation of an integrated quantum optics controlled-NOT chip. Single photons (represented as flashes) propagate on the chip, confined by silica waveguides, and are then coupled into optical fibers for detection. See page 646.

Image: W. Amery/Bristol University

DEPARTMENTS

EDITORIAL

NEWS OF THE WEEK

NEWS FOCUS

600

LETTERS

BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

620

Science

SCIENCE EXPRESS
www.scienceexpress.org

TECHNICAL COMMENT ABSTRACTS

REVIEW

623 & 643

BREVIA

RESEARCH ARTICLES

REPORTS

Science

CONTENTS

REPORTS *CONTINUED...*

664

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

Seeing the light.

SCIENCE NOW
www.sciencenow.org
HIGHLIGHTS FROM OUR DAILY NEWS COVERAGE

Gene Therapy Improves Sight of Four Patients
Treatment could potentially be applied to host of blindness disorders.

The Oceans Finally Show Their Stripes
High-resolution ocean map reveals a worldwide pattern of striated currents.

No Substitute for Real Blood
Analysis links blood-replacement products to higher risk of heart attack and death.

Exercise, a natural antidepressant.

SCIENCE SIGNALING
www.sciencesignaling.org
THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

PERSPECTIVE: VGF, a New Player in Antidepressant Action?
J. E. Malberg and L. M. Monteggia
The neuropeptide VGF appears to play a role in the antidepressant effects of exercise, the neurotrophic factor BDNF, and the neurotransmitter serotonin.

PERSPECTIVE: Synaptic Patterning by Morphogen Signaling
W. R. Williamson and P. R. Hiesinger
The morphogen Activin acts as a permissive and local motility restriction signal around individual photoreceptor terminals.

PODCAST
E. M. Adler and A. M. VanHook
Some glial cells can generate action potentials and are hypersensitive to ischemic injury.

New structure for university careers in Europe.

SCIENCE CAREERS
www.sciencecareers.org/career_development
FREE CAREER RESOURCES FOR SCIENTISTS

Structuring Academic Careers in Europe
C. Wald
Several European universities have begun defining tenure-like career structures.

Taken for Granted: Lost in Space
B. L. Benderly
A severe disconnect separates some policy-makers from scientists' real lives.

Mastering Your Ph.D.: Exploring Nonprofit Organizations
P. Gosling
The values and culture of nonprofit organizations make them an exciting and rewarding career choice.

May 2008 Funding News
J. Fernández
Learn about the latest research funding opportunities, scholarships, fellowships, and internships.

SCIENCE PODCAST

Download the 2 May *Science* Podcast to hear about a neural circuit for babbling in young songbirds, reversing loss of sight, lung scans for cancer detection, and more.
www.sciencemag.org/about/podcast.dtl

Separate individual or institutional subscriptions to these products may be required for full-text access.
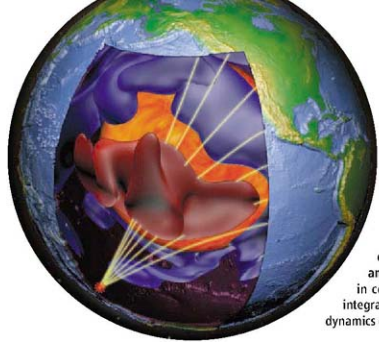
<< Journey to the Center of the Earth

Our view of Earth's lowermost mantle has changed recently because of the likely presence of a newly identified mineral phase, post-perovskite. Furthermore, recent advances have improved seismic imaging of the structure and composition of this region. **Garnero and McNamara** (p. 626) review these developments, and, in combination with recent modeling efforts, provide an integrated view of the region and how it may influence mantle dynamics overall.

Improving Thermoelectrics by Powdering and Pressing

Thermoelectric materials are semiconductors that combine high electrical and low thermal conductivity to allow the recovery of electrical power from waste heat or direct cooling with applied current. One route to improving the figure of merit that describes these materials, ZT, is to form them as nanocrystalline materials, because the added grain boundaries should help to scatter phonons and lower their thermal conductivity. **Poudel et al.** (p. 634, published online 20 March) take one of the most widely used thermoelectric materials, p-type $Bi_xSb_yTe_3$, which has ZT of 1 at room temperature, and convert it to a nanocrystalline form. They ball-milled the material into nanoparticles under an inert atmosphere, and then hot-pressed the powder back into bulk ingots. This material has a peak ZT of 1.4 near 100°C, and could produce temperature differentials of 100°C in a cooling mode.

Shake Unrattled by Roll

When molecules are excited by absorption of light, there is a brief period during which quantum mechanical oscillations retain a specific phase relationship and so can be manipulated productively. However, perturbations due to the onset of various vibrations and rotations, as well as the influence of neighboring molecules, soon randomize this coherent state and inhibit further active control. **Branderhorst et al.** (p. 638) explored the tendency of a vibrational mode in an ensemble of potassium dimers to lose coherence due to mixing with molecular rotation. As a marker for coherence, they detected the persistence of quantum beats, a result of wave interference in the fluorescence signal. By iteratively shaping their laser excitation pulse using this marker as feedback, they succeeded in prolonging coherence by a factor of 2 and accounting for the efficacy of the pulse shape with a model.

From the Mouths of Baby Birds

The youthful noises of young zebra finches sound different from adult zebra finches, somewhat like babbling in human infants. Using surgical and pharmacological lesions, **Aronov et al.** (p. 630) eliminate some of the brain regions and neural connections that support adult song. The lesions cause the adults to

sound again like juveniles, but leave juvenile vocalizations intact. Thus, the brain connections upon which bird song depends differ between adults and juveniles, and the process of song maturation is not simply a refinement of an existing neural network, but involves switching from a youthful network to one required for adult song.
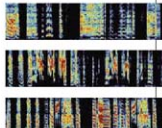
Quantum Optical Chips

While the quantum mechanical aspects of optics have been illustrated in a number of applications in communication, metrology, and lithography, these have generally been carried out on a room-sized optical bench. Quantum technologies-based photonics will require scaling down and operating on a platform that is robust and easy to implement. Patterning optical waveguides and circuits using silica-on-silicon, **Politi et al.** (p. 646, published online 27 March; cover) demonstrate that arbitrary photonic quantum circuits can be realized on silicon chips. Single photons launched into the devices showed high visibility interference and basic quantum logic operations with high fidelity.

Prostratin Preparation

Despite progress in drug development to target HIV, the virus remains very hard to root out of an infected system entirely, because of latent reservoirs beyond the reach of current treatments. Certain compounds, chief among them prostratin, have recently shown promise for accelerating emergence from these reservoirs, which may enhance the long-term effectiveness of other drugs. However, the scarcity of botanical sources for prostratin has hampered progress. **Wender et al.** (p. 649) present an efficient four-step chemical synthesis to produce prostratin from a much more abundant natural precursor. Moreover, the route can easily be modified to afford structural analogs that may enhance therapeutic efficacy.

Mitochondria as Drivers of Metastasis

Most cancer deaths occur when cells in a primary tumor metastasize, yet the mechanisms by which tumor cells acquire metastatic properties remain poorly understood. **Ishikawa et al.** (p. 661, published online 3 April) explored the role of mitochondria in this process by taking mouse tumor cell lines with either a high or low propensity to metastasize and swapping their mitochondrial DNA (mtDNA). Interestingly, the recipient cells acquired the metastatic potential of the cells donating the mtDNA. In one tumor cell line examined in detail, the mtDNA conferring high metastatic potential was found to harbor mutations that led to overproduction of reactive oxygen

species (ROS), and up-regulation of nuclear genes involved in metastasis. Pretreatment of tumor cells with ROS scavengers reduced their ability to metastasize in mouse models, suggesting a possible avenue for the development of therapies to suppress metastasis.

Sugars in Living Color

The glycan structures that adorn the cell surface are a rich source of information about biochemical activities occurring inside the cell. Levels, distribution patterns, and structural changes of sugars making up the glycans reflect modifications in flux through metabolic pathways, alterations in gene expression, and changes in secretory pathway dynamics. The direct and noninvasive visualization of glycans in vivo has proved difficult. Now **Laughlin** *et al.* (p. 664) report the in vivo multicolor, time-resolved imaging of cell surface glycans in live, developing zebrafish. Dramatic bursts in glycan production were observed in the jaw, olfactory organ, and pectoral fin during the period of 60 to 72 hours post-fertilization, with major tissue-specific differences in the levels and trafficking patterns of glycans during embryogenesis.

Expanding Low O$_2$ Zones

One of the consequences of a warming climate, warmer oceans, is expected to cause a decrease in the oxygen concentration of the oceans. This prediction is based on the fact that the solubility of oxygen decreases as water temperature increases, as well as the modeled result of a slower rate of advection of water to the deep ocean while sinking organic matter continues to decay in a process that consumes oxygen. **Stramma** *et al.* (p. 655) report measurements of dissolved oxygen concentrations in the tropical Atlantic and equatorial Pacific oceans that show a clear vertical expansion of the oxygen minimum zones over the past 50 years. A reduction of the concentration of dissolved oxygen could have serious effects on marine life, especially in regions that are already at the limits of oxygen concentration required to support many organisms.
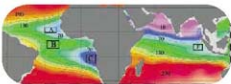
Testing Tethers

The Golgi complex is composed of a set of flattened membrane cisternae interconnected by trafficking membrane vesicles. Inside the cell, during transport through the Golgi complex, membrane vesicles seem to rely on membrane tethers to avoid their escape from the Golgi region. Such membrane tethering has seldom been reconstituted using pure proteins and artificial membranes. **Drin** *et al.* (p. 670) present and test a simple model for how a long coiled-coil protein of the Golgin family, GMAP-210, forms a bridge between a highly curved membrane (a vesicle) and a flat one (a cisterna). This asymmetric tethering relies on motifs that sense membrane curvature. The tethering mechanism presented is asymmetric and reversible, which may explain how the Golgi keeps a constant morphology despite being constantly remodeled by membrane transport.

A Sense of Danger in the Air

Particulate airborne pollutants, such as asbestos and silica, are notorious for their negative effects on health, including lung inflammation and cancer, yet information on how such substances exert their effects is lacking. **Dostert** *et al.* (p. 674, published online 10 April; see the Perspective by **O'Neill**) reveal that a multiprotein complex known as the Nalp3 inflammasome can signal exposure of cells to internalized particles of asbestos and silica, which leads to the activation of a potent inflammatory response. In the absence of Nalp3, mice responded less vigorously to asbestos, supporting the idea that this inflammatory sensing complex plays a key role in the response to respiratory pollutants.

Unintentional Uptake

In many respects, our understanding of innate immune responses to protozoan parasites still lags behind that for other infectious organisms. However, recent work has shown that an important part of the armory against African trypanosomes is serum apolipoprotein L-I (apoL1), which can kill the parasite by causing lysis—why then would the parasites take it in? **Vanhollebeke** *et al.* (p. 677) show that apoL1 is taken up by the parasite via a specific glycoprotein receptor, which the parasite normally uses to supply heme for its growth and resistance to oxidative stress within the host. In human serum, however, the receptor also inadvertently recognizes a component of certain high-density lipoprotein complexes, of which apoL1 is a part, explaining how the uptake of this detrimental host protein is triggered.

CREDIT: STRAMMA ET AL.

Donald Kennedy is Editor
Emeritus of *Science*.

Misbegotten Preemptions

The notion of preemption has a long history in relation to the U.S. Food and Drug Administration (FDA). Its primary significance had to do with the Commerce clause in the Constitution, which (along with the Supremacy clause) gives the federal government power to regulate commerce between the states. When I was commissioner of the FDA in the late 1970s, my colleagues and I rather liked preemption. Suppose, for example, a state decided to set its own net weight requirements for packaged foods so as to favor its own manufacturers. Well, just because it wanted to disfavor out-of-state competition, it wouldn't be allowed to. Similarly, if it wanted to establish its own drug approval agency, that would also be preempted by the FDA's authority.

Of course, there are contemporary reasons for being less enthusiastic about this kind of regulatory preemption. Today it is being used to prevent states from undertaking actions to protect their environments when they are especially vulnerable to certain insults. Regulatory preemption, for example, has killed carbon-sparing gas mileage provisions undertaken by the state of California because it has particular problems with the Los Angeles airshed. And California can't even regulate soot produced by vessels off its own shore because the federal Clean Air Act preempts that, too.

But the very notion of preemption has taken on an entirely new guise, also involving the FDA, but in a far more troubling way than the older use of preemption in the regulatory sense. This radical change comes from the Bush Administration's chief counsel at the FDA, Daniel Troy. His private career, before his government appointment in 2001, included membership on the Legal Policy Advisory Board of the conservative Washington Legal Foundation. During his first year, Troy developed a reputation for having his door open to industry for private discussions, the notes on which could not be made public. After returning to private practice, he published a piece in *Legal Times* entitled "When the FDA Acts, State Torts Must Defer." The title gives a clear message: If the FDA has approved a drug or device, the manufacturer is immune from product liability lawsuits.

This odd concept, gaining favor in some state courts, is not only bad policy; it could be dangerous to your health. Why? First, the FDA is badly underfunded. Recent flat budgets have hurt the agency, and despite efforts by FDA advocates, the outlook is grim. Congress has relied too much on the Prescription Drugs User Fee Act, first passed in 1992. Unfortunately, most of that user fee money can only be used in the process for approving new drugs; only a trivial fraction can be used to strengthen safety monitoring of already-approved drugs.

Second, the nature of the FDA's standard process makes it unable to make a secure guarantee of safety. Approval of a drug for a given indication follows a series of controlled clinical trials. But even for a drug expected to have millions of potential users, the experimental limb of the trial (in which participants receive the drug rather than a placebo) will have only a few hundred to a thousand patients. Once the drug is in wide distribution, it may have a thousand times as many users. It's no surprise that widely marketed drugs produce scary media accounts involving a threatening adverse reaction that appears suddenly, resulting in deaths or serious illness, and ending in withdrawal from the market or strengthened label warnings.

The FDA cannot claim infallibility in its premarketing procedures. Nor is the postmarketing safety monitoring system adequate. It depends on voluntary reporting of adverse reactions by doctors, and because there is no way of knowing how many patients are taking a particular drug, the rate of an adverse reaction cannot be determined. In view of these deficiencies, how can one seriously defend a no-liability clause to protect the manufacturer? In short, if you can't sue the maker of a product, you deserve some guarantee that it's safe. If the FDA can't provide that, why should you and I find the courtroom door closed?

– Donald Kennedy

CHEMISTRY
Chloramine Complexities

Chloramine is a comparatively recent weapon in the ongoing battle to eliminate harmful microorganisms from drinking water supplies. Though its disinfecting properties are straightforward, the concomitant generation of ammonia as a byproduct can give rise to a complex web of downstream chemistry that remains an active area of study. One important reaction is microbial nitrification, or oxidation of the ammonia to nitrite and nitrate, which also lowers the water's pH by acid production. Zhang et al. have systematically explored the efficiency of nitrification in plumbing pipes of differing compositions—polyvinyl chloride (PVC), copper, lead, and brass—at various pH and phosphate levels. They found that relative to PVC, copper inhibited nitrifier growth, whereas lead enhanced it (probably through reductive cycling of nitrate back to ammonia via lead corrosion). Brass initially resisted nitrification activity, but then shifted its behavior after ~120 days, as the efficiency of copper leaching from the alloy diminished. A perhaps counterintuitive consequence of this reaction web is that PVC pipes may ultimately cause more metal ion leaching into the water stream than copper pipes, as the acid byproducts of nitrification degrade brass valves and faucets. — JSY

Environ. Sci. Technol. **42**, 10.1021/es702483d (2008).

ECOLOGY
Frogs Leap to Extinction

The causes of recently documented declines in frogs since the 1980s have been hotly debated. One vigorously promulgated hypothesis is that the decline has been triggered by climate change, which has promoted virulence in a previously saprophytic fungus. An orthogonal view is that the decline reflects the spatiotemporal spread of an invasive fungal disease. In either scenario, the fungus is *Batrachochytrium dendrobatidis*, which colonizes frog skin and suffocates the amphibians. The declines have been particularly noticeable among the charismatic harlequin frogs of Central and South America. Lips et al. have developed a technique to analyze the unavoidably incomplete frog census data (due to infrequent sampling, remote habitats, or sociopolitical challenges) and see wavelike progressions of population falloffs that look very much like the spread of an invasive pathogen originating from three source locales. They categorically found no relation with climate change; indeed, the fungus does best at altitudes where conditions are cool and moist. — CA

PLoS Biol. **6**, e72 (2008).

MATERIALS SCIENCE
An Extended Jog

Radiation striking crystalline materials can damage their structure and related properties by generating vacancies (missing atoms) or intersti-

tials (extra atoms stuffed between lattice sites). Demkowicz et al. use simulations to probe the effects of adding or removing atoms in copper-niobium multilayer nanocomposites. Two flaw-free interfaces can form between the Cu and Nb. The first occurs from the joining of the face-centered cubic Cu [111] plane and body-centered cubic Nb [110] plane. The second requires a straining and rotating of the Cu [111] to make it about 0.5% less dense in its interfacial area (Cu$^\alpha$). Under strain, screw and edge dislocations can form in the various layers, but of particular note, the screw dislocations can sit either at the Cu-Nb interface, or can shift into a Cu-Cu$^\alpha$ interface. Thus, there are pathways for defects to move from the Nb into the Cu, for example. When an atom is removed from or added to the Cu layer, the authors found that instead of generating a localized defect, the atoms would reconstruct to form an extended jog (shown at left) that interacted and annihilated with existing screw dislocations. The efficient defect recombination suggests that materials designed to have similar interfaces to the Cu-Nb system could be useful for limiting damage from radiation exposure. — MSL

Phys. Rev. Lett. **100**, 136102 (2008).

MOLECULAR BIOLOGY
Not an Open and Shut Case

Eukaryotic cells have evolved a complex machinery to ensure a precise and equal segregation of their chromosomes during cell division. At the center of this machine is the kinetochore, a large multiprotein complex found in the centromeric region of each chromosome. Kinetochores bind to the microtubules that pull replicated chromosomes apart bodily, giving one each to the daughter cells. Centromeres, and kinetochores too, are specified epigenetically—that is, not directly from signals in the underlying DNA. In order to manipulate the epigenetic state of kinetochores, Nakano et al. have constructed a human artificial chromosome (HAC) bearing a kinetochore with a permissive protein-binding site at its heart. These artificial kinetochores mimic the behavior of their natural counterparts, but they are completely disrupted—and the artificial chromosome is lost from the cell—when a protein that silences transcription binds to them.

The silencing protein nucleates the formation of repressive (or closed) heterochromatin, and it is this epigenetic change that inactivates the kinetochore. Surprisingly, the binding of an activating protein at the same site also interfered with HAC segregation, suggesting that kinetochore function is highly sensitive to the architecture of the chromatin in which it is embedded. — GR

Dev. Cell **14**, 507 (2008).

ECOLOGY

Fire in the Far North

Paleoecological data sets contain historical records of biotic responses to changes in climate. Currently, high-latitude regions are suffering a particularly aggressive regimen of climate change; hence, an understanding of past vegetation dynamics in these regions is especially pertinent. Higuera *et al.* have analyzed pollen records from north-central Alaska and find that a combination of drier climates and shrubbier tundra during the late glacial period 14,000 to 10,000 years ago led to regular fires. Given present-day increases in shrub biomass and temperature, tundra fire activity might increase again, with consequences for vegetation dynamics and carbon cycling. Tinner *et al.* have analyzed pollen and other records from the past 700 years (a period that includes the Little Ice Age of 1500 to 1800 CE) in southern Alaska, and find that temperature fluctuations of $1°$ to $2°C$, together with changes in moisture balance, led to conversions between boreal forest and tundra with concomitant alterations in fire regimes. Taken together, these findings are consistent with models predicting a conversion of tundra to boreal forest as temperatures increase. — AMS

PLoS ONE **3**, e0001744 (2008); *Ecology* **89**, 729 (2008).

PHYSICS

Heralded Photon Purification

Quantum information processing requires techniques for the generation and transportation of quantum data. The use of photons as the carriers of that quantum information, with the data encoded as different polarization states of the light, is particularly appealing because photons are robust against decoherence effects and can be transported over long distances. The photons, however, need to be indistinguishable; i.e., identical and in a pure quantum state. Although techniques exist for the generation for single photons, determining whether they are in fact indistinguishable and pure has generally required a postselection or spectral filtering process that compromises their quantum utility. By careful design of the parametric downconversion process, in which a single photon is divided into two entangled photons, Mosley *et al.* show that restricting the optical modes into which the photon pairs emerge can provide a method for the generation of heralded (by measurement of one member of the pair) indistinguishable single photons of high purity. — ISO

Phys. Rev. Lett. **100**, 133601 (2008).

Science Signaling 〈〈 **Just When You Thought It Was Pseudo…**

Approximately 10% of the known protein kinases are thought to be catalytically inactive, and therefore dubbed pseudokinases, because they lack one or more conserved motifs in their active sites. The pseudokinase Ca^{2+}/calmodulin (CaM)–activated serine-threonine kinase (CASK) has an altered DFG motif, which would normally bind a Mg^{2+} ion that coordinates the phosphoryl group to be transferred from ATP onto the substrate. CASK is known to bind to synaptic adhesion molecules, including neurexin, and *CASK*-deficient mice exhibit synaptic defects and perinatal death. Mukherjee *et al.* determined the structure of the CaM-kinase domain of CASK and found that it resembles a constitutively active kinase. They also show that a fluorescent ATP analog bound to recombinant CASK in the absence of Mg^{2+} and that adding Mg^{2+} inhibited this interaction. In vitro assays revealed that the CASK CaM-kinase domain exhibited autophosphorylation activity and that Mg^{2+} and other divalent cations inhibited this activity. Finally, overexpression of wild-type CASK in rat hippocampal neurons resulted in increased phosphorylation of neurexin, challenging the idea that pseudokinases act merely as inactive scaffold proteins. — JFF

Cell **133**, 328 (2008).

CREDIT: WILLY TINKER

Chicken Raised in a Dish

Research prizes are all the rage, but a new one is sure to raise eyebrows if not gorges: $1 million for getting "in vitro" (IV) meat onto supermarket shelves. People for the Ethical Treatment of Animals (PETA) announced the money pot last week, acknowledging on its Web site that although it would prefer that consumers stick to vegetables, "many people continue to refuse to kick their meat addictions." IV meat, the thinking goes, would at least save a lot of animals.

So far, a few scientists have been trying to get cells from pigs and other animals to grow in the lab. But big challenges remain, such as finding the optimal culture conditions. Last month, the first international IV meat symposium was held in Norway.

PETA is asking contestants to submit IV chicken samples by June 2012 to its panel of 10 judges; entrants must also be able to mass-produce the meat. Although $1 million is a nice bonus, "the real prize would be the global meat market, which is worth hundreds of billions of dollars annually," assuming consumers are willing to eat the stuff, says Jason Matheny, founder of New Harvest, a nonprofit that promotes substitute meat.

Bird and Mammal Make a Couple

Researchers have witnessed a wild Antarctic fur seal attempting sex with a king penguin—the first documented case of a pinniped trying to mate with an animal that is not only a different species but also a different class of vertebrates.

P. J. Nico de Bruyn and colleagues at the University of Pretoria in South Africa saw the 45-minute attack at a beach on Marion Island in the southern Indian Ocean. "The seal alternated between resting on the penguin and bouts of pelvic thrusting copulatory behavior," the team writes in the May Journal of Ethology. Burney le Boeuf, a behavioral ecologist at the University of California, Santa Cruz, says he's not surprised, given the male propensity to father as

many offspring as possible: "Sperm are cheap." Le Boeuf says he's reminded of male wild turkeys, which will attempt to mate with a stick if it's placed at an angle that mimics the neck of a receptive female.

Glyphs for Docs

Physicians may be losing their attention spans along with the rest of us, but French researchers have come up with a remedy: pictographs to give doctors shortcuts to information that will help them write better prescriptions.

Jean-Baptiste Lamy, a bioinformatics expert at the University of Paris, and colleagues call their new language VCM for *visualisation des connaissances médicales*. Use of the system will avoid a lot of common prescribing errors, they argue, as doctors often don't have time to read drug monographs when making medical decisions. "A VCM-based software can help [a physician] verify that a drug can be prescribed to the patient without contraindication," says Lamy. "It has been shown that this step is sometimes skipped due to the lack of time."

In the current issue of the journal BMC Medical Informatics and Decision Making, the authors report that they successfully trained 11 general practitioners in their system of icons that represent symptoms, diseases, drugs, and tests, and can be combined for more complex meanings. The symbols above, for example, substitute for "The hypokalemia caused by this drug increases the cardiac toxicity of digitalis glycosides and the risk of heart rhythm disorders."

EARLY OILS

Buddhist artists in Bamiyan, Afghanistan, may have painted with oils centuries before European Renaissance painters developed the technique.

A team led by Marine Cotte at the European Synchrotron Radiation Facility in Grenoble, France, has analyzed tiny samples of paintings sent by a UNESCO conservation team from a site where the Taliban destroyed two giant Buddha statues in 2001. Initial scans with ultraviolet light led researchers to suspect the presence of oil, and "we have confirmed it," says Cotte. Twelve of 50 murals depicting colorful Buddhas and mythical creatures, painted in caves behind the statue niches, included pigments bound in plant oils. Oil offers "more freedom" to artists, says Cotte, as it doesn't set instantly like the gypsum or calcium salt pigments also used in the caves.

Helen Howard of the National Gallery in London says European oil paintings date back to the 12th century, but whether oil was used earlier isn't known because "analysis hasn't often been carried out on very early paintings." UNESCO team leader Yoko Taniguchi of the National Research Institute for Cultural Properties in Tokyo said in a statement that ancient Romans and Egyptians were known to use drying oils, but only as medicines and cosmetics. Thus, the team writes in April's Journal of Analytical Atomic Spectrometry, the Afghan samples could be the "oldest example of oil paintings on Earth."

Celebrities

KARL MARX PLANCK? Did Max Planck, the founder of quantum physics, share a name with the intellectual father of communism? In preparation for the celebrations surrounding the physicist's 150th birthday on 23 April, a German television reporter searching old church records for the names of Planck's godparents found more than he was looking for. The entry in the baptismal registry reads "Karl Ernst Ludwig Marx Planck, goes by Marx."

It's not clear whether the pastor improperly recorded the name or whether Planck later changed it to avoid confusion with the political philosopher, who was 39 and widely known when Planck was born. Lorenz Beck of the Max Planck Society's Archive in Berlin votes for the former, based on a letter from 10-year-old Planck, which he clearly signed Max. But any conclusive proof may have gone up in flames when a 1944 bombing raid destroyed Planck's house in Berlin. Bernd Wirsing, spokesperson for Germany's Max Planck Society, harbors no doubts. "It's worth a footnote for historians," he says. "But we don't know of any document where he called himself Marx. He was always Max."

MONEY MATTERS

BIG BUCKS. A biotech company formed by Harvard University antiaging researcher **David Sinclair** is being acquired by pharmaceutical giant GlaxoSmithKline for about $720 million. Glaxo is paying nearly double the share price, or $22.50 a share.

Sinclair started Sirtris Pharmaceuticals to develop resveratrol, a molecule in red wine, into a drug for various age-related diseases (*Science*, 27 February 2004, p. 1276). The 4-year-old company, based in Cambridge, Massachusetts, is testing a more stable version of resveratrol, dubbed SRT501, in people with diabetes and a rare mitochondrial disorder and studying small molecules that activate the same gene, $SIRT1$, but at a much lower concentration.

According to SEC filings, Sinclair owns 153,000 shares in Sirtris, worth more than $3.4 million. He says he's bet "my whole reputation" on the company. "Neither my wife nor I are that focused on the money," he says. Sirtris will be an autonomous unit within Glaxo.

IN BRIEF

Steven Kurtz, an art professor at the University at Buffalo, New York, who in 2004 was indicted for receiving bacterial cultures by mail, was cleared by a federal judge last week. Kurtz was planning to use the microbes in an art exhibit (*Science*, 9 July 2004, p. 159). University of Pittsburgh geneticist Robert Ferrell, who bought the microbes and shipped them to Kurtz, was fined $500 and sentenced to a year of unsupervised probation in a ruling on the case earlier this year.

Economist **W. Brian Arthur** of the Santa Fe Institute in New Mexico and mathematician **Yakov Sinai** of Princeton University are the inaugural winners of the Lagrange Prize for research on the science of complexity. They will receive $118,000 each from Italy's CRT Foundation.

Got a tip for this page? E-mail people@aaas.org

THEY SAID IT

"Religious belief is not science. Science and religious belief are surely reconcilable, but they are not the same thing."

—Raymund Paredes, Texas Higher Education Commissioner, explaining last week why he had turned down a certification request from the Institute for Creation Research for a Master of Science degree in science education.

Awards >>

BARREN TO LUSH. Martin Fisher has won a $100,000 Award for Sustainability from the Lemelson-MIT program for inventing manual pumps to help small farmers in Africa.

Fisher, a mechanical engineer, began work on technologies for the rural poor after a trip to Peru in 1984. His Super MoneyMaker pump draws water from as deep as 7 meters and pushes it uphill with pistons powered by stair-climber–like treadles. It can irrigate 0.8 hectares with less than a day's labor and minimal environmental impact. The sturdy metal pump, which costs about $100, typically pays for itself within 6 months, Fisher says, and can increase a farmer's profits 10-fold. More than 95,000 pumps have been sold since 1996, mainly in Kenya, Tanzania, and Mali.

Fisher emphasizes that the key to the pump's success is that local companies produce and sell it. He runs a social entrepreneurship nonprofit, Kickstart, founded in Kenya and headquartered in San Francisco, California. He'll share part of the prize with his business partner, Nick Moon, and plow the rest into cheaper, more powerful pumps.

NEWS>>

THIS WEEK

GINA's champion
596

Fossil food webs
598

ENDANGERED SPECIES

Spotted Owl Recovery Plan Flawed, Review Panel Finds

A blue-ribbon panel of scientists has confirmed major flaws in the proposed recovery plan for the northern spotted owl, a threatened species that has driven forest policy in the northwestern United States for nearly 2 decades. As did earlier reviews, the final one, by the Sustainable Ecosystems Institute (SEI) in Portland, Oregon, concludes that the Fish and Wildlife Service's (FWS's) plan does not put enough emphasis on protecting the owl's habitat. It also says that massive thinning of dry forests is needed to prevent habitat from going up in smoke—a recommendation that makes some environmentalists nervous. All eyes are now on FWS, which intends to release a final version of the plan by the end of the month. The plan "could have a very profound effect," says forest ecologist Jerry Franklin of the University of Washington, Seattle, who participated in the SEI review.

The northern spotted owl (*Strix occidentalis caurina*) lives mainly in old-growth forests, which range from British Columbia to northern California. After decades of logging and population decline, the species was put on the federal list of threatened species in 1990. FWS began to create a recovery plan to use in determining habitat critical for the species' survival. Although recovery plans lack the legal clout of regulations, they are a major influence on FWS decisions to permit logging and other activities that might harm a particular species. The recovery plan for the spotted owl was never finished, however, because it was superseded by the overarching Northwest Forest Plan, which guides forest policy across the region.

Many environmentalists think the Northwest Forest Plan wasn't protective

In trouble. The 45-cm-tall northern spotted owl is threatened by logging of old-growth forests, fires, and invasion of its habitat by barred owls.

enough—the spotted owl population continues to decline by 3.7% a year—whereas the timber industry has complained that it's too restrictive (*Science*, 29 July 2005, p. 688). In 2002, the American Forest Resources Council (AFRC), an industry group, sued FWS for failing to complete a review of the owl's status, which it is required to do every 5 years. As part of a settlement, FWS agreed to reexamine its designation of critical habitat by 1 June 2008. A separate lawsuit also led the agency to finalize its recovery plan for the species.

In spring 2006, FWS formed a team to draft the recovery plan that included a broad range of expertise, including environmentalists and timber industry representatives, but

lacked top scientists; some declined to participate in part because they feared the process would be politically charged, they told *Science*. The team's draft focused on the need to protect habitat and also dealt with the threat from barred owls, an invasive species that is competing with the spotted owl.

Politics did trump science, say observers and participants. After the first draft was sent to Washington, D.C., in September 2006, officials at the Department of the Interior (DOI) ordered the recovery team to add another management strategy, called Option 2, says recovery team member Dominick DellaSala, an ecologist who directs the National Center for Conservation Science & Policy in Ashland, Oregon. This option would reduce the amount of land set aside for owl conservation and give the Bureau of Land Management (BLM) and the Forest Service more flexibility to allow logging. The immediate goal was to make the recovery plan consistent with a BLM proposal to facilitate logging in Oregon, according to internal agency e-mails provided to *Science* by DellaSala. He says that officials also wanted the plan to list the barred owl threat as more dire than loss of habitat—over the objections of some of the recovery team members, as well as James Tate, DOI's own science adviser.

FWS released the draft plan, including Option 2, for public comment and requested scientific review in April 2007. Anonymous peer reviews, organized by the Society for Conservation Biology and two other science groups, raised many concerns. In August, for example, reviewers recruited by the Wildlife Society, a nonprofit association of wildlife conservation and management experts, called the draft plan "seriously flawed." The reviewers deemed the draft's overall focus on barred owls rather than habitat conservation "incredibly risky." And Option 2, they concluded, "drastically reduces protection for owl habitat and maximizes flexibility given to land managers by allowing them to operate under a series of nebulous rules." The Wildlife Society urged the agency to start over.

In December 2007, the agency contracted with SEI to analyze all the reviews and suggest scientifically valid recovery options. In a 157-page report released on 21 April 2008, nine leading owl and forestry experts echoed many of the previous criticisms of the draft plan. They confirmed that barred owls are a

FOCUS

Smoker
screens
600

Promising
gene therapy
606
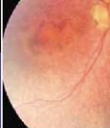
threat, recommended experiments to determine how they can be controlled, and reiterated that the draft plan underestimated the importance of protecting habitat.

In a major departure from both previous reviews and the draft plan, the SEI panel called for much more aggressive thinning to reduce the risk of massive forest fires, especially in the dry, eastern part of the spotted owl's range. "We think the threat of wildfire is so great that we need to do thinning," says

lead author Steven Courtney of SEI. Ecological restoration is also necessary, Franklin adds. In contrast, DellaSala and reviewers for the Wildlife Society say that more needs to be learned about possible detrimental effects of thinning on spotted owls.

The recovery plan is now being finalized to meet the June deadline for revising critical habitat, says FWS spokesperson Joan Jewett. She expects that thinning will be addressed. The timber industry agrees that thinning, and

barred owls, are a serious threat, says AFRC President Thomas Partin. But he dismisses the impact of logging large trees, because he says that has been relatively minimal in recent years. DellaSala counters that the science says every hectare of owl habitat matters. Given the pressure from BLM and the Forest Service, he's pessimistic about how much protection the final plan will afford old-growth forests. "It might get decided in the courts," he notes.
–ERIK STOKSTAD

GLOBAL WARMING

Mother Nature Cools the Greenhouse, But Hotter Times Still Lie Ahead

As climate-change skeptics like to point out, worldwide temperatures haven't risen much in the past decade. If global warming is such hot stuff, they ask, why hasn't it soared beyond the El Niño–driven global warmth of 1998? Mainstream climate researchers reply that greenhouse warming isn't the only factor at work. And in a new paper, they put some numbers on that rebuttal. They show that regional and even global temperatures are being held down by a natural jostling of the climate system, driven in large part by vacillating ocean currents. The study "shows how natural climate variability can mask the global warming effect of greenhouse gases," says climate researcher Adam Scaife of the Hadley Centre for Cli-

mate Prediction and Research in Exeter, U.K., "but only for a few years."

The latest reminder of climate's confounding subtleties comes in climate forecasts that Noel Keenlyside of the Leibniz Institute of Marine Sciences in Kiel, Germany, and colleagues published this week in *Nature*. Rather than simply predicting temperatures at the end of the century, as most modelers do, they ran their simulations only 10 and 20 years into the future. At such a time range, short-term effects can override the contributions of rising greenhouse gases (*Science*, 10 August 2007, p. 746). For example, great, heat-carrying currents like the Gulf Stream can slow down and speed up, cooling and warming surrounding conti-

nents. As a result, temperatures rise and fall from decade to decade even in the absence of human interference.

To take account of such ocean-driven natural variability, Keenlyside and his colleagues began their model's forecasting runs by giving the model's oceans the actual sea surface temperatures measured in the starting year of a simulation. Providing the initial state of the ocean doesn't make much difference when forecasting out a century, so long-range forecasters don't usually bother. But an initial state gives the model a starting point from which to calculate what the oceans will be doing a decade hence and therefore what future natural variability might be like.

The added observations did in fact improve simulations of past climate variations. Looking into the future, the model forecasts a slowing of heat-carrying Atlantic currents and thus a cooling over the North Atlantic, North America, and western Europe in the next decade. It even predicts a slight cooling of the globe. But by 2030, forecast global temperatures bounce back up to the warming predicted with greenhouse gases alone.

The forecast is not the first to herald a slowing or even a temporary cessation of global warming. A study involving even more ocean observations inserted at the beginning of model runs reached similar conclusions last year. "The different approaches give slightly different results," says climate modeler Douglas Smith of the Hadley Centre, who headed the earlier study, but "we do agree there's a temporary offset of global warming due to natural variability." So if you're a climate-change activist pointing to year after year of mounting climate crises, you might want to rethink your approach.
–RICHARD A. KERR
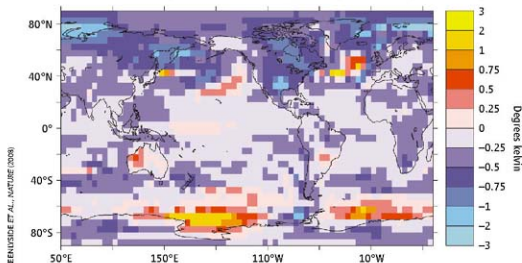
Not so hot. In a model's simulation of climate out to 2015, much of the world is cooler (blues) than it would have been without natural climate variations driven by vacillating ocean flows.

CREDIT: KEENLYSIDE ET AL., *NATURE* (2008)

Long haul. Representative Louise Slaughter (D–NY) has pushed a ban on DNA-based discrimination through 13 years of congressional debate.

LOUISE SLAUGHTER INTERVIEW

How to Get a Genetic Protection Law Through Congress? Keep Trying

Persistence pays. After 13 years of rejection, a bill to protect individuals from employment and insurance discrimination based on their DNA, the Genetic Information Nondiscrimination Act (GINA), is poised to become law. Last week, the U.S. Senate voted unanimously to approve the bill; the House of Representatives was scheduled to vote on it as *Science* went to press. And President George W. Bush has vowed to sign it if it reaches his desk.

GINA has had many champions, but one of its most stalwart backers and the only co-author who trained in microbiology and public health is Representative Louise Slaughter (D–NY). Slaughter, 78, has introduced the bill in every Congress since 1995. She was first driven by concerns that individuals would be reluctant to participate in genetic research without the protections that GINA offers. But the changing genetic landscape has led her to expand her thinking, and she now hopes that GINA will put families more at ease about taking one of the growing number of genetic tests hitting the market.

Navigating the politics that held GINA up has been tricky. The bill didn't make it to a vote in the House until after the Democrats took control of Congress in 2006 and overcame the concerns of health insurers and other businesses that feared it

would raise costs. GINA passed twice in the Senate, in 2003 and 2005, but was later blocked single-handedly by Senator Tom Coburn (R–OK). Coburn, who said the bill might expose businesses to legal liabilities, recently withdrew his objections. Slaughter was upbeat in a conversation with *Science* earlier this week.

–JENNIFER COUZIN

Q: Why did you first introduce the bill back in 1995?
L.S.: I'm a scientist, and I believed in the Human Genome Project. We didn't know much about it at the time, but I really felt that since I was in Congress ... that the social policy should keep up with the science. I wanted them to move in tandem.

Q: Why has it been such a long road?
L.S.: I think it's been a total misunderstanding. ... In the Senate, a single man held up 85 bills including this one—he *didn't* understand it.

Q: You're talking about Senator Tom Coburn?
L.S.: Yes. [For many years,] this bill was held up in the House; there was never a hearing. We always had far more cosponsors than we needed to pass it but were not able to get it on the calendar.

Q: Why was that?
L.S.: The drug companies and the insurance companies didn't want it.

Q: Why not?
L.S.: They were mistaken. Everyone has bad genes. Dr. Francis Collins [head of the U.S. National Human Genome Research Institute in Bethesda, Maryland, and a longtime champion of GINA] had said to us early on, the president of the drug company and the president of the insurance company have bad genes. It's to their benefit and the benefit of their families that we know all we can.

Q: And you were thinking about this particularly in terms of research?
L.S.: It started in research, but we knew the promise of this science was so great, being able to provide health care in an entirely new way. Not only to save billions of dollars in health care money, but we could save in human suffering. It never made any sense to me that we would have all these roadblocks.

Q: How will GINA help?
L.S.: People will not be afraid to allow themselves to be part of a research program because they will know it will not affect their employment or insurance. ...

But the most important thing is the number of people [who worried about this]. ...We had one constituent who had a lot of Alzheimer's in his family. He very much wanted to know if he had that gene in his family but was afraid to get tested. Many went to clinics [for genetic testing] under assumed names.

Q: Are more people going to buy genetic tests now? Is that problematic, given that there are concerns about the quality of these tests?
L.S.: We may get higher quality now. Once we got the bill passed, obviously protecting your information means getting a good test as well, in my book.

Q: What's next for you in this area?
L.S.: I'm in a wonderful spot for a microbiologist—given food that's unsanitary, outbreaks of *E. coli*. There's a number of things that we are very much aware of and worry about. We want the FDA [U.S. Food and Drug Administration] to be much stronger, much, much, much stronger.

ECOLOGY

Yosemite: Protected but Not Preserved

The spectacular landscape of California's Yosemite Valley looks natural and has been protected for more than a century. Yet ecologists know that today's valley is not the one that enchanted naturalist John Muir when he promoted the formation of Yosemite National Park. One of the park's emblematic species, the California black oak (*Quercus kelloggii*), is in decline, a problem usually attributed to a lack of fire and an invasion of conifers. Now, two researchers have traced the trees' suffering through a complex chain of effects that starts with human influence and extends from cougars through mule deer to oaks and primroses. The result of that "trophic cascade" is that there are very few young oak trees to replace their elders, hurting other species from shrubs to birds and invertebrates and apparently reducing overall biodiversity, according to a report in the May issue of *Biological Conservation*.

"It's not just Yosemite," says ecologist Thomas Rooney of Wright State University in Dayton, Ohio, who was not involved in the study. He notes that other oak species are in decline across the United States. "It shows that habitat protection alone is not enough. You need the predators."

In Yosemite, the paper says, the direct cause of oak mortality is the high density of mule deer, which have been munching the oaks' basal sprouts and seedlings for nearly a century. Intriguingly, the mule deer's abundance stems not from a lack of predators—as with elk in Yellowstone National Park (*Science*, 27 July 2007, p. 438)—but from shy ones: elusive cougars (*Puma concolor*). The mountain lions keep the deer in check elsewhere in the park but avoid areas like the valley, where people congregate.

"There are higher deer densities now than in the 1850s," when American settlers first entered Yosemite Valley, says ecologist William Ripple of Oregon State University, Corvallis, who co-authored the study with OSU colleague Robert Beschta, a forest hydrologist.

After Yosemite became a national park in 1890, visitors swarmed into the valley, eager to see its parklike landscape of black oaks and famed wildflower-filled meadows. Officials began eliminating cougars, bobcats, and coyotes and protecting mule deer from hunting. By 1925, deer were

numerous, and park observers noted a marked decline in the most popular wildflower, the evening primrose (*Oenothera hookeri*). Although not apparent at the time, that's also when the black oak seedlings began to disappear, say Ripple and Beschta, who in 2006 measured the diameters of more than 3000 black oaks. They also took tree ring cores from 40 sites close to the valley's visitor center and sites 4 to 8 kilometers away. Black oaks can live as long as 500 years, but a "healthy stand includes a mix of young and old trees," says Beschta. Oaks close to the visitor center have almost no young trees or basal sprouts. The deer are using people as "protective shields," says Ripple. Deer had also nipped off the flower buds of nearly every evening primrose the scientists saw.

Free lunch. Mule deer near Yosemite's visitor center feast on oak seedlings and evening primrose flowers (inset).

Today, Yosemite managers burn prescribed areas to keep out conifers and clear the way for oaks. But fires can't do what most needs to be done, says Ripple: "Get the baby oaks to grow." That's not likely to happen until there are fewer deer. "It wouldn't be popular to have culling in a national park," Rooney says, "but it may be necessary" if the valley's biodiversity is to be preserved.   **–VIRGINIA MORELL**

CREDITS: W. J. RIPPLE/OREGON STATE UNIVERSITY

Cancer Genome Goes Global

There's an ambitious new sequencing project on the block: the International Cancer Genome Consortium (www.icgc.org). Leaders aim to raise $1 billion to sequence 50 human cancers over the next 10 years and share the data. This week, it joined a crowded field; similar efforts are under way at the U.K. Sanger Institute and the U.S. National Institutes of Health (*Science*, 8 September 2006, p. 1370). But a global organization makes sense because the prevalence and environmental causes of cancer differ around the world, says consortium leader Thomas Hudson of the Ontario Institute for Cancer Research in Toronto: "We're trying to prepare ourselves for the next wave." Organizations in nine countries, including in China, Singapore, and India, have signed on.   **–JOCELYN KAISER**

Stresses Grow in U.K. Science

U.K. parliamentarians attacked the Labour government this week for slighting science and mismanaging the current allocation of £2.8 billion. The science committee in the House of Commons also leveled harsh words at the agency that supports astronomy, particle physics, and government labs, saying it had axed fields and facilities without consulting the community and citing "particular weakness" in its peer-review systems and management. Neglect has "caused immense damage to fundamental science in this country," says particle physicist Brian Cox of the University of Manchester. However, U.K. innovation secretary John Denham argued in a speech that "as a government, we have fought for, and won, record resources" for science.
                                    –DANIEL CLERY

Wage Understanding, Not War

The social and behavioral sciences may get as much as 20% of a proposed $250 million boost to the U.S. Department of Defense's basic research budget to counter terrorist threats without force. "We have given our troops many technologies to win conflicts, but we haven't done enough to help them avoid conflict," William Rees, the Pentagon's chief of basic research, told *Science* last week. Rees was amplifying a message from other officials, including Defense Secretary Robert Gates, who credits a small team of anthropologists embedded with military units in Afghanistan for helping to reduce violence in the region.
                          –YUDHIJIT BHATTACHARJEE

TROPICAL DISEASES

Dispute Clouds the Future of U.S. Naval Lab in Indonesia

A United States naval laboratory in Jakarta that studies tropical diseases may fall victim to Indonesia's determination to develop its own research capabilities and take control of its H5N1 viral samples. Negotiations over the lab's status have dragged on for more than 2 years, and now some Indonesian politicians are calling for it to be closed or taken over by Indonesia. Recently, the Indonesia Ministry of Health ordered institutions to stop sharing all medical samples with the facility, which has severely limited what it can do.

Opened in 1970, the U.S. Naval Medical Research Unit No. 2 (NAMRU-2) is one of five U.S. tropical disease research labs scattered around the world. NAMRU-2's main function is to study infectious diseases that might affect U.S. troops, although it has taken on a wider public health role in working with Indonesia's Ministry of Health, for instance, in supporting efforts to help curb malaria and dengue. In addition, NAMRU-2's staff of

about 20 Americans and 150 Indonesians has trained hundreds of Indonesian health workers, researchers, and students in lab techniques. It also monitors emerging diseases throughout Southeast Asia in collaboration with local institutes.

Responding to e-mailed questions, Trevor Jones, NAMRU-2's commanding officer, wrote, "The U.S. benefits because American scientists have the opportunity to study tropical disease transmission where it is actually occurring." He added that the restrictions have slowed the lab's work, but he hopes this is temporary.

One of the most worrisome diseases circulating in the world is the H5N1 strain of avian influenza, and Indonesia is by far the hardest-hit country, with 132 con-

In the spotlight. Officials opened the U.S. Naval Medical Research Unit No. 2 to the public in Jakarta to show that most employees are Indonesians.

firmed human cases and 107 deaths. NAMRU-2 performed the lab diagnosis of all human cases of H5N1 in Indonesia from June 2005 to January 2007, when Indonesia's Ministry of Health labs took over the duties. Since then, Indonesia has only sporadically shared samples of the H5N1 virus with the World Health Organi-

PALEOECOLOGY

Fossils Help Figure Out Food Webs Old and New

The watery world 540 million years ago abounded with exotic life forms rivaling those created by Dr. Seuss. Ecologists have long wondered how these worms, seaweeds, sponges, trilobites, snails, and meter-long beasts with rings of teeth interacted. Now, a daring analysis of fossils from China and Canada shows that these marine plants and animals from the Cambrian Period formed food webs on par with those existing today. These ancient networks follow the same rules seen among inhabitants of current reefs, deserts, and bays, report Jennifer Dunne, an ecologist at the Santa Fe Institute in New Mexico, and colleagues online 28 April in *PLoS Biology*.

"I was surprised that something like this [study] is really possible," says Volkmar Wolters, an ecologist at Justus Liebig University in Giessen, Germany. "The result is a thought-provoking, highly informative, and breathtaking paper on the potential structure of Cambrian communities."

Food webs are the complex networks of feeding interactions among the plants, animals, and microbes in a particular ecosys-

tem. Over the past 30 years, researchers have learned that food webs share certain patterns, irrespective of the habitat or the particular species involved. The number of species making up the web and their degree of connectedness dictate certain elements—for example, the number of top predators, the percentage of omnivores, and so on. Dunne wondered how early these patterns emerged, but the fossil record had seemed too sketchy to provide detailed enough information about who eats whom.

Douglas Erwin, a paleontologist at the Smithsonian National Museum of Natural History in Washington, D.C., knew of two exceptions. After a lunchtime conversation with Dunne, he offered to help compile the necessary data from two Cambrian fossil beds, the Burgess Shale in Canada and the Chengjiang Shale in China. At both sites, soft as well as hard body parts were preserved. He and Rachel Wood of the University of Edinburgh, U.K., combed the literature for descriptions of gut contents, bite marks, teeth, and other structures indicative of particular eating habits. They

established links among herbivores, predators, and prey for 142 plants and animals in the Burgess Shale and 85 in the Chengjiang formation and rated how confident they were of each link.

Dunne and her colleagues Richard Williams of Microsoft Research Limited in Cambridge, U.K., and Neo Martinez of the Pacific Ecoinformatics and Computational Ecology Lab in Berkeley, California, used these data to compare the two Cambrian ecosystems to eight modern ones, including a reef, an island, and a pond. They calculated the total number of links for each species in each ecosystem and the nature of those connections, for instance, how many were predator-prey, or herbivore-plant. They plugged the number of species and links into a mathematical model that describes modern feeding systems to see if it could accurately predict what the Cambrian food webs would look like.

It showed that the food-web structure—the numbers of organisms at each level, for example—was quite similar in all ecosystems studied. And it also showed that in some ways,

zation (WHO), demanding more equitable access to any vaccines or other benefits derived from those samples, which are used to monitor virus evolution, drug resistance, and pandemic risk (*Science*, 23 February 2007, p. 1065).

Now that dispute has extended to NAMRU-2 and is hampering its research on all tropical diseases. When the lab's Memorandum of Understanding expired in December 2005, U.S. officials assumed negotiating a new agreement would be routine. NAMRU-2 continued its work until the Ministry of Health halted all sample sharing on 31 March. Tensions escalated further after an early April visit by U.S. Health and Human Services Secretary Michael Leavitt, who wrote on his blog on 14 April that Indonesia's Minister of Health, Siti Fadilah Supari, "has used the sample-sharing debate and the negotiations over the status of NAMRU-2 in Indonesia to set herself up as an antagonist of the United States." Last week, several Indonesian politicians joined the fray, calling for NAMRU-2 to be closed or taken over by Indonesia.

The U.S. Embassy and NAMRU-2 are now trying to reassure Indonesians of the lab's good intentions and negotiate a new agreement. John Heffern, deputy U.S. ambassador, held a press conference on 24 April to defend the lab's activities, and NAMRU-2 held a media open house on 25 April. One remaining point of contention, Heffern told the press, is providing diplomatic immunity for the Americans working at the lab, which the U.S. believes to be standard practice but many in Indonesia are now questioning. Local newspapers have reported that other issues pertain to technology transfers. (Calls seeking comment from Indonesia's Ministry of Foreign Affairs were not returned.)

The outcome of the negotiations will also affect NAMRU-2's status as a WHO collaborating center for emerging diseases. That designation, in effect since 1997, is on hold pending the resolution of its status. David Heymann, who heads WHO's pandemic influenza efforts, says the lab has been "very important" but adds that Indonesia's own lab capabilities are advancing rapidly. "We are encouraging developing country labs to become collaborating centers," he says.

Jones declined to comment on the negotiations. But he emphasized that "we do not have plans to move to another Asian country. We are dedicated to a future here in Indonesia."

–DENNIS NORMILE

---

the Chengjiang food web—15 million years older than Burgess—was more primitive. The researchers found more "loops" in the Chengjiang web, wherein the same species appears in a particular food chain. By contrast, the Burgess Shale and modern food webs tend to be more hierarchical, a trait considered important for stability, Dunne notes. Another analysis revealed that the Chengjiang food web was more loosely connected than the rest. Today, any species in a web is so closely connected to others that a change in one tends to affect most of the web members. In China, that may have not been the case.

But 15 million years made a difference. "The younger Burgess Shale web looks incredibly like the modern food webs," Dunne points out.

The work is "an excellent contribution to both paleoecology and food-web theory, show-

On the menu. Ball-and-stick diagram shows who, such as the predatory *Anomalocaris*, eats whom in this 500-million-year-old food web.

ing the relevance of the fossil record to understanding current ecosystem states," says Peter Roopnarine, a paleontologist at the California Academy of Sciences in San Francisco.

However, he and Wolters wonder whether Dunne and colleagues were overly bold in assuming fossils at a particular locale really coexisted, as the beds cover millions of years, and whether the sampling was comprehensive enough for this sort of analysis. Just because the structure of the food webs seems similar doesn't mean they functioned the same way, cautions Roopnarine, who says that the paper "falls short on some of its claims." Nonetheless, he thinks the work will have an impact: "The questions emerging from this paper should encourage paleontologists to think more seriously about the need to develop theoretical and modeling approaches to fossil ecologies."

–ELIZABETH PENNISI

Business Boost Thwarted

A congressional Democrat with clout and a Republican with conviction have teamed up to block a plan to give small businesses a bigger slice of the federal research pie. Last week's vote by the House of Representatives came on a bill to reauthorize two research programs that fund peer-reviewed proposals from start-up companies through a tax on 11 science agencies. Of greatest concern to science lobbyists was language raising the share going to the SBIR (Small Business Innovation Research) and STTR (technology transfer research) programs from a combined 2.8% to 3.6%, an increase that would have siphoned off an additional $650 million a year. But representatives David Obey (D–WI), chair of the powerful appropriations committee, and Vernon Ehlers (R–MI), a former physics professor who had failed to derail the increase during an earlier committee vote, argued successfully on the House floor that this was the wrong time to tap already stressed science budgets.

A larger SBIR program "does no harm for a large agency whose budget has been rising, such as the Department of Defense," Obey said shortly before last week's vote, "but it can do immeasurable harm to the crown jewel of our research agencies in this country, the National Institutes of Health." The White House also opposed the increase. A proposal for an even larger boost has stalled in the Senate. Both programs are set to expire this fall unless Congress reauthorizes them.

–JEFFREY MERVIS

Campaign Bailout for Arecibo?

Senator Hillary Clinton (D–NY) has introduced legislation (S. 2862) to keep federal funds flowing to the Arecibo Observatory in Puerto Rico. Her support for the world's largest single-dish radio telescope, which is slated to lose National Science Foundation support by 2011 (*Science*, 21 September 2007, p. 1663), would benefit both her home state of New York—the observatory is operated by Cornell University—and the economy of Puerto Rico, which holds a presidential primary on 7 June. Clinton, who trails Barack Obama in the race for the Democratic nomination, hailed Arecibo's "remarkable tools for understanding the universe and the "path-blazing accomplishments of these New Yorkers." Puerto Rico's delegate introduced a similar bill last fall in the House of Representatives. A Cornell spokesperson said the university was "absolutely pleased" by Clinton's move.

–ANDREW LAWLER

A Bruising Battle Over Lung Scans

Doctors and researchers are sharply divided over the merits of screening smokers and others at high risk of lung cancer with costly CT scans; a $200 million clinical trial has become a lightning rod

Sheila Ross is known as a "two-time survivor" at the advocacy group where she works. Doctors found cancer in her lung, treated it, and declared the treatment a success—twice. "I am very lucky" to be alive, Ross acknowledges, because the odds of surviving 5 years after one such diagnosis are only about 15%. She lost a lung in the second treatment, and at that point, after 20 years as a congressional aide, she says, "I knew exactly what I had to do." She joined a Washington, D.C., activist group, the Lung Cancer Alliance. Its style, like Ross's, is impatient.

The alliance kicked a hornets' nest last year, criticizing the cancer-research establishment for passivity and demanding that the U.S. government investigate two clinical leaders for alleged financial conflicts (see sidebar, p. 602). That was merely a side skirmish in a battle over early detection of lung cancer, a fight that has split the cancer-research and treatment communities. At issue is whether an advanced x-ray imaging technique known as low-dose spiral computed tomography (CT) scanning should be widely used to catch tumors when they're small, in hope of removing them before they spread. The alliance, along with an assertive group of scientists, argues that CT's effectiveness is already established and that it should be widely used to screen those who have a high risk of lung cancer, primarily smokers and former smokers.

An equally impassioned group of scientists argues that without better information, such a step would be folly. They say the benefits of CT screening for lung cancer aren't proven and that its widespread use could do more harm than good. It would be expensive, too, adding billions of dollars to the annual cost of Medicare and private insurance. Some argue that the government would get more reliable results by investing in tough measures to curb smoking, which is the biggest cause of lung cancer. "Essentially no one believes the current data support the hypothesis that screening is beneficial," says outspoken skeptic Peter Bach, a lung-cancer specialist at the Memorial Sloan-Kettering Cancer Center in New York City.

Caught in the middle of this dispute is the National Cancer Institute (NCI). It is investing in a huge clinical trial, the largest cancer screening test it has ever run, in the hope of getting some guidance. The first data are due by 2010, although it may take longer to sift out key results. The trial itself has come under fire from the alliance, which wants to cut short what it considers an agonizing wait for action. The advocacy group has appealed to the U.S. Congress and government agencies to bypass the academic debate and move straight to screening. On the other side, the skeptics say that although the trial may not be perfect, it must be completed before a decision on screening is made.

A lot rides on the outcome. Lung cancer kills more than 160,000 people a year in the United States, and if more tumors can be caught early, many lives could be saved. But expanding CT screening would subject a large swath of the public each year to ionizing radiation, biopsies, and surgery; it would cause an unknown number of deaths itself. Every smoker in the United States is a

High definition. Computerized reconstruction of x-ray images makes it possible to detect very small nodules in the lung.

prime candidate for lung imaging, and there are an estimated 45 million of them.

Catching small tumors

Even people who disagree on screening agree that CT images are a vast improvement over the chest x-ray. Engineers and computational wizards came together in the mid-1990s to create machines that could probe large areas in the body noninvasively and with precision. In a CT scanner, the x-ray source and detectors spin in a ring around the patient, who slides past the moving beam to yield a spiral "slice" through the body. A CT scan can be quick—5 seconds—and uses less radiation than traditional methods. Speed also makes it possible to get an image in a single breath, reducing blur from motion. And unlike an old chest x-ray, the data can be tweaked by computer to tune certain densities in or out and reconstruct three-dimensional views of tissue and bone.

Starting in 1998, the race to upgrade imaging machines "really took off," says Peter Kingma, U.S. vice president for CT imaging at Siemens, the German manufacturer. During a period known as the "slice wars," companies increased the number of detectors rapidly; the capability to produce two simultaneous helical slices rose to 16 per scan when the big NCI screening trial began in 2002 and has continued climbing to 256 slices per scan and higher. Resolution also jumped from 10 millimeters in the early days to 0.625 mm; Kingma says

low-dose images now resolve an area about 0.3 mm square.

Perhaps the most controversial pioneers in CT scanning for diagnosis of lung cancer are Claudia Henschke and David Yankelevitz, both of the Weill Cornell Medical College in New York City. In a landmark paper in the 26 October 2006 issue of *The New England Journal of Medicine* (*NEJM*), Henschke, Yankelevitz, and a group of international partners—known as the International Early Lung Cancer Action Project (I-ELCAP)—reported a stunning success: 88% of the early stage lung-cancer patients they detected lived at least a decade after diagnosis.

More than 40 participating centers in I-ELCAP adopted a method first developed on a small scale by Henschke and Yankelevitz in New York. They recruited subjects at somewhat elevated risk of lung cancer and gave them all CT scans. In I-ELCAP, those without tumors were asked to return for a scan in a year. The rest went on to a series of branching paths: If imaging found a specific solid nodule with a diameter of at least 5 mm, or a nonsolid one of 8 mm diameter, they went under intensive study. Some got a biopsy and then possibly surgery or other treatment. Others, if their tumor looked less threatening based on texture or growth rate, were asked to come back in 3 months. The aim in the I-ELCAP studies is to continu-

ously sift the population, identify fast-growing cancers, and remove them.

The group led by Henschke claims that with experience it has grown better and better at distinguishing slow-growing from fast-growing tumors. This minimizes unnecessary biopsies, the clinicians say,

> *"Essentially no one believes the current data support the hypothesis that screening is beneficial."*
> —PETER BACH, MEMORIAL SLOAN-KETTERING CANCER CENTER

but enables them to intervene early when a tumor changes. In the *NEJM* article, I-ELCAP reported that of the 31,567 people age 40 or older who initially enrolled in the study, 4186 (13%) were flagged by baseline CT scans as positive and required follow-up. But only 484 eventually received a diagnosis of lung cancer; 412 of these had the earliest form, stage I. It was in this latter group that the estimated 10-year survival rate was 88%. The authors say that only by providing good follow-up care and tracking people for more than 5 years can you see the full benefits of screening. Some studies that report no benefit didn't do aggressive follow-up or observe patients long enough, they argue.

The results had an impact in a field desperate for good news. "For the first time in my professional career, here is something that is offering to mitigate the largest cause of cancer [mortality]," says David Burns, a lung-cancer expert and professor of family medicine at the University of California, San Diego, who has a long career battling tobacco and defending federal antismoking measures. Now he is defending CT screening. "The logic is difficult to argue with," he says: "You can clearly find [lung cancer] smaller, and you can clearly find it at an earlier stage. If you don't treat it, people die." So why not try to make CT screening work?

Doubts

Hope and enthusiasm are important, but "science has to be cold-eyed," says Bach of Sloan-Kettering. He thinks the I-ELCAP screening study is riddled with flaws. He's also upset by the publicity it's received, which he worries may have increased the number of people seeking CT scans. The procedure is "incredibly lucrative," he notes. The charge for an

CT champion. Claudia Henschke's group at Weill Cornell Medical College has reported that vigilant screening catches lung tumors early and saves lives.

A Bumper Crop of Conflicts

A clash between clinical researchers over whether former smokers and others at high risk for lung cancer should be screened using computed tomography (CT) scans (see main text) has turned bitterly personal. Some of the most contentious questions have been about intellectual and financial conflicts.

The Lung Cancer Alliance (LCA), a patient advocacy group in Washington, D.C., cast the first allegation. LCA's president, Laurie Fenton Ambrose, is irate that the U.S. government has refused to endorse CT imaging for lung-cancer screening while it awaits results from a $200 million trial to evaluate the procedure, the National Lung Screening Trial (NLST). She has charged that some leaders of the trial revealed their bias against CT screening when they agreed in past years to testify for tobacco companies about how screening might do more harm than good. Ambrose and an ally at another advocacy group leveled these charges in a blitz of correspondence to federal agencies, targeting two distinguished NLST leaders—radiologists William Black of Dartmouth Medical School and Denise Aberle of the University of California, Los Angeles.

The letters prompted several inquiries, including one in the U.S. House of Representatives last fall led by Michigan Democrat John Dingell.* Dingell's probe, which made headlines and then faded from view, is "active and ongoing," committee staff claim.

Black and Aberle have acknowledged that they agreed to testify for tobacco companies but said they did nothing improper. Aberle, who coordi-

* energycommerce.house.gov/Investigations/NIH.101907.NIH.NCI.ltr.pdf

nates a large network of NLST clinical centers, provided testimony in 2003 in a class-action trial in Louisiana. As Aberle explained in a letter to the National Cancer Institute, she "violated no conflict of interest disclosure requirements," and the checks she received—reportedly totaling about $30,000—went to her university. In the letter, she said she wanted to "articulate the uncertainties of CT screening and the potential risk" to people in Louisiana who might sign up for it. Black similarly agreed to provide testimony in 2006 for attorneys defending Philip Morris in a New York class-action suit by smokers who wanted the company to pay for their annual CT scans. As Dartmouth's general counsel explained in a letter to Ambrose, Black believes that widespread screening may "cause more harm than benefit" and prepared testimony about why it would be a mistake for the court to set a precedent for screening. But he changed his mind, withdrew, and returned a $700 payment because he realized his participation "might be misconstrued as support for the tobacco industry." A review by the National Institutes of Health found that neither grantee had violated rules on disclosing conflicts of interest.

As the dust settled on this controversy, *The Cancer Letter*, a Washington, D.C., weekly, published an exposé of potential conflicts on the other side of the debate. It revealed that two well-known researchers who claim unprecedented success with CT screening for lung cancer—Claudia Henschke and David Yankelevitz, both of the Weill Cornell Medical College in New York City—have a financial stake in an invention that could be used in connection with CT screening. They have applied for 27 patents related to lung screening and have accepted royalty income from one license, but, *The Cancer Letter*

---

initial CT scan may be modest—about $200 to $300—but that's just the first installment. An anomaly "gins up all kinds of business," says Bach, as clinicians follow up with positron emission tomography scans, biopsies, and other tests. Imaging is a gateway into high-cost medicine and has been flagged as a growing budget concern by the U.S. Medicare program.

The glaring weakness of I-ELCAP, according to Bach, who with epidemiologist Colin Begg of Sloan-Kettering and others published a study on lung screening last year in *The Journal of the American Medical Association*, is that it is not a randomized controlled trial. This makes it susceptible to bias. The best way to avoid bias in a screening trial, they argue, is to randomly assign patients to receive a CT scan or no CT scan and then keep track of who dies.

Without random selection, trial results can be dramatically skewed, for example, by "lead-time bias." It produces the illusion that early diagnosis is responsible for extending the life of a patient when in fact the patient has just received a diagnosis earlier.

Other common problems, called "length bias" and "overdiagnosis," arise from the imprecision of cancer biology. Too little is known about early stage tumors to predict which will become malignant; intensive screening can flag many that are benign or slow-growing as dangerous when they really

are not. "Pseudodisease" is the term used by William Black, a radiologist and lung-cancer specialist at Dartmouth Medical School, to

Resolving power. Although new imaging techniques (*above*) offer more information than the chest x-ray (*right*), they also deliver more false-positive signals.

describe this byproduct of screening. He and his Dartmouth colleague, clinical epidemiologist H. Gilbert Welch, argue that this is a big medical risk that clinicians need to guard against. In addition to causing harm, overdiagnosis can boost the number of people who are diagnosed with cancer and appear to overcome it. "Everyone

should know that when you go down this road [of cancer screening]," says Welch, "there is going to be harm; the question is, what will the benefit be."

The study by Bach and others—a comparison of a validated model of clinical experience with data on 3246 patients from three CT screening trials—found "no evidence" that screening reduced the risk of death from lung cancer in a period of almost 5 years. But screening dramatically boosted medical workups. The authors found that biopsies increased threefold above the expected level; lung surgeries, 10-fold.

Henschke and Yankelevitz claim that the extreme vigilance built into their approach keeps overdiagnosis and other biases to a minimum; clinicians intervene if "a malignant rate" of growth is evident. In addition, she and Yankelevitz write in the January 2008 issue of *The Oncologist*, a panel of pathology experts has examined all specimens removed by surgery and "confirmed that they are all genuine lung cancers and that 95% of them are already invasive."

Tobacco's dividend? CT screening to catch lung cancer early is being considered for all smokers—and there are 45 million of them in the United States.

charged, they did not properly disclose these interests in medical journal articles. In addition, *The New York Times* and *The Cancer Letter* reported in coordinated articles that most of the funds supporting the Weill project came from a tobacco company gift of $3.6 million.

Henschke and Yankelevitz have since acknowledged that their widely cited 2006 article in *The New England Journal of Medicine*, for one, should have disclosed that they received royalties from their patented "methods to assess tumor growth and regression in imaging tests"—inventions that have been licensed to General Electric (GE), a maker of CT machines. In addition, they acknowledged that "virtually all" of the money from a foundation listed as a sponsor of their research actually came from an "unrestricted gift by the Vector Group, the parent company of Liggett Tobacco, which manufactures cigarettes." In a separate statement, Weill says that Vector's original pledge was disclosed and reported in the national press 5 years ago and should be viewed in the same light as funding that "peer institutions and medical schools" received from antitobacco lawsuits.

Even the group that first raised these questions may have a conflict of its own. Ambrose acknowledges that LCA, a tireless advocate for government action to expand CT imaging, has received funding from GE. Ambrose says the alliance always made known that it receives 40% of its funding from "corporate interests," including the unrestricted GE grant and a larger one from a biotech company involved in lung-cancer research.　　　　**–E.M.**

None of this satisfies the skeptics. Bach's doubts have grown so that he now says: "We worry that the basic principle [of CT screening] is wrong. … Most of the lung cancers that are claiming lives, we think, are coming like a meteor. They come out of nowhere and are everywhere." Screening can't catch them. Yet others argue that Bach has gone overboard. Says James Mulshine, a leader of the Lung Cancer Alliance and associate provost for research at the Rush University Medical Center in Chicago, Illinois: "I haven't seen evidence in the literature that supports" Bach's view of meteorlike cancers.

Bruce Chabner, editor-in-chief of *The Oncologist* and clinical director of the Massachusetts General Hospital Cancer Center in Boston, says he's planning to air new concerns that go beyond study design in an editorial about the I-ELCAP results. For example, he claims that, unlike all clinical trials sponsored by drug companies and NCI, this privately funded project has not submitted its data to an outside audit. The Weill researchers did not respond to a request for comment.

A hard endpoint

NCI's proposed answer to the confusion is to look for help from a $200 million project it is now funding, the National Lung Screening Trial (NLST), a randomized controlled study. From 2002 to 2004, it enrolled and screened more than 50,000 individuals through a network of more than 30 study sites in the United States. The volunteers, all with an elevated risk for lung cancer, were randomly assigned to receive a chest x-ray or CT scan. Individual centers have been following up with standard monitoring and therapy. From 2008 on, researchers will be adding up deaths until they detect a statistically valid result showing that more people died in the x-ray group or the CT group—or neither.

By 2010, the first results should be available from NLST. But CT screening advocates have already been taking shots at it. For example, some suggest that it was a mistake—perhaps unethical—to recruit people with the promise of high-quality diagnosis and then give chest x-rays, long viewed as a poor diagnostic tool. Henschke and Yankelevitz stopped using chest x-rays early in their study because, as they wrote in *The Oncologist*, it "missed" 76% of the screening-diagnosed cancers found by CT.

The Lung Cancer Alliance also questioned whether patient follow-up was aggressive enough throughout the NLST network, because a slow response could make the diagnostic method look poor. NCI Director John Niederhuber responded in a letter last year that treatment "is not standardized in the NLST." But he argued that this should not compromise the trial because "variations in treatment should occur equally in both arms." According to Laurie Fenton Ambrose, president of the Lung Cancer Alliance, the emphasis on counting deaths rather than aggressively screening and treating patients is akin to "doing nothing" and is "just not acceptable."

Last year, Ambrose and other leaders of the pro-screening movement appealed to NCI for an interim view of CT screening, before NLST is done. They proposed combining data from I-ELCAP with data from NCI-funded trials, including NLST and another known as PLCO, in an attempt to get an early sense of the potential value of CT screening. Niederhuber met with the petitioners but decided it would not be "appropriate or fiscally responsible" for NCI to hold a review, he wrote to Ambrose.

Otis Brawley, chief medical officer of the American Cancer Society (ACS), has agreed to serve as a broker. He is not an advocate of trying to get an early view of CT screening's benefits. (The idea was proposed by an epidemiologist at ACS, Robert Smith.) But Brawley says that he intends to host a meeting of experts on the topic; NCI and major international cancer institutions will be invited to participate. Brawley aims to bring investigators together in May or early June from four randomized trials, including three from Europe, and, "perhaps" someone to represent the I-ELCAP study. It will be a kind of "grand jury," he says, to review the trials and see whether it would be possible to use existing data to conduct a meta-analysis of CT screening.

This grand jury may not lead to a new course of action, but it could help bring some calm to a hotly contested field of clinical research.　　　　**–ELIOT MARSHALL**

Wishing for an African Einstein

Hoping to nurture Africa's talent into becoming elite scientists, cosmologist Neil Turok has plans to clone the math institute he founded
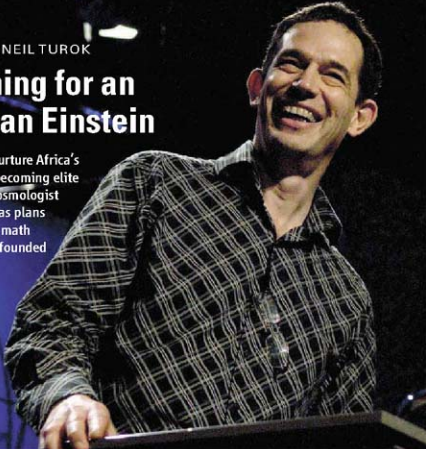
CAMBRIDGE, U.K.—In 2001, mathematical physicist Neil Turok went back for the first time in 25 years to his childhood home, South Africa, to visit his parents. Dismayed by the lack of opportunities for math graduates in Africa and motivated by his father, a former antiapartheid activist, the University of Cambridge researcher took action. Over the next 2 years, Turok had a derelict building near Cape Town renovated into a new institute, enrolled 29 math graduates from 11 African nations, and persuaded mathematical colleagues to teach there for 3-week shifts. "It's a very inspirational venture, … a real flagship project," says Britain's Astronomer Royal Martin Rees, who has visited the institute.

The African Institute for Mathematical Sciences (AIMS) continues to grow (see sidebar, p. 605), but Turok isn't stopping there. He's leading the effort to replicate it at 15 centers across the continent, each focusing on a different area of applied math, such as economics. "When people hear about AIMS, they get very excited," Turok says, "and people see the spark of something much, much bigger."

Turok's dream has gotten a massive boost from TED, a conference held every year in California, with about 50 invited speakers—originally just from the fields of technology, entertainment, and design, hence TED. Three of the annual speakers receive a $100,000 prize, and their talk must contain "one wish to change the world." Last year, biologist E. O. Wilson wished for an online encyclopedia of life; it debuted a few months ago. This year, Turok was one of the chosen.

A prizewinner's talk at the TED conference is essentially a sales pitch to philanthropists and companies; Wilson raised $50 million after his plea. Turok says he knew instantly what his wish would be: for the next Einstein to be African. Worrying that his Cambridge colleagues would ridicule such a notion, Turok says he tried it out on a few particularly hard-nosed individuals and didn't get one negative comment.

Something is needed to jump-start African science, most agree. There have been many well-intentioned efforts to boost the continent's universities, but outside South Africa, world-class research still struggles to find a home and few pursue science professionally (*Science*, 9 February 2007, p. 748). "Africa has 922 million inhabitants, more than 40% of whom are under age 14. There may well be an Einstein here, but we must find better ways to tap into that vast pool of young talent," says Sudanese mathematician Mohamed Hassan, president of the African Academy of Sciences and executive director of TWAS, the Academy of Sciences for the Developing World. Rees agrees: "It's important for Africa to develop its own intellectual elite and develop the conditions to keep them there."

Turok started off in Africa. He was born in 1958 in Johannesburg, where his parents, members of the African National Congress,

> "When people hear about AIMS, they get very excited, and people see the spark of something much bigger."
>
> —NEIL TUROK

fought apartheid. They were imprisoned and finally expelled from South Africa; Turok spent much of his childhood in Tanzania and Kenya before settling in London for high school. At Cambridge University, he says he "fell for theoretical physics." After completing his doctorate, he worked at two U.S. universities and the Fermi National Accelerator Laboratory, returning to Cambridge in 1997.

During this time, he carved a name for himself devising observational tests for current cosmological theories. In recent years, he's developed, in collaboration with Princeton University theorist Paul Steinhardt, a model for a cyclic universe in which the big bang is the result of two multidimensional membranes, known as brane worlds, colliding. In this model, the big bang is just the start of the latest phase of a universe that explodes, expands, contracts, implodes, and then starts all over again.

Turok's own life split into two parallel universes during that 2001 visit to South Africa. He calculated that there must be thousands of math graduates emerging from African universities every year. "Most are not able to find work and are frustrated because they can't do the interesting stuff," he says. Turok's father, who was by then a member of South Africa's parliament, provided "a huge stimulus" for AIMS. He refused to let his son sit down and watch an important rugby match until he'd put his idea for a mathematics institute down on paper; the father then promptly faxed the document to his political friends. "He forced me into it by pure embarrassment," recalls Turok.

The project won support from the universities of Cambridge, Oxford, and Paris-Sud XI, and the three universities of the Western Cape. The South African government and charitable foundations also pitched in; educating a graduate at AIMS costs roughly a fifth as much as sending them to Europe or the United States.

Turok's experience with AIMS so far has confirmed his belief that the talent is there in Africa. It just needs a way to break out, he says. Most African math or science graduates would be overwhelmed, contends Turok, if they went straight to a European or North American research university. Having a postgrad center in Africa, he says, lets students "gain confidence. It changes their whole perspective."

By 2005, African politicians had taken notice of the AIMS model, and a plan was hatched to replicate it in universities across Africa. Under the auspices of the African

An African Showcase for Math Studies

MUIZENBERG, SOUTH AFRICA—An exotic mixture of languages—Swahili, Amharic, Malagasy—echoes through the lobby of this former beach hotel that has been transformed into a hothouse for cultivating bright young African minds. But it is the universal language of mathematics that unites the diverse and ambitious group of students making a second home here.

The 53 students at the African Institute for Mathematical Sciences (AIMS)—whether they come from the deserts of northern Africa, such as Esra Khaleel of the conflict-torn Darfur region, or from the lush southern island of Madagascar, like Mika Sidonie Ranaivomanana—share a passion for numerical concepts and a determination to make a difference in their home countries. "While I am at AIMS, my goal is to understand the difficult concepts," says Proscovia Namayanja of Uganda. "I want to return to Uganda to teach those concepts." AIMS's director, theoretical physicist Fritz Hahne, former dean of science at the University of Stellenbosch in South Africa, describes his charges as "creative and committed students" who "are here because they want to *give* something to Africa and to their home countries."

AIMS alumni surveys indicate that the vast majority of graduates go on to study for advanced degrees, mostly at South African universities but also in Europe and North America. It is unclear exactly how many return to their native lands, although every one of the dozen current students interviewed by *Science* said that they eventually would. Walter Mudzimbabwe of Zimbabwe, for example, plans to become an expert in a field that might benefit his hyperinflation-plagued homeland: financial mathematics. On the steps outside the AIMS building, Lydia Flore Mamoade—the first woman to attain a mathematics degree in her country—discusses her plans to earn a higher degree and then return home to the isolated Central African Republic.

AIMS accepts only about one in five applicants. Students, who are given free room and board, take a series of intensive 3-week courses from visiting lecturers, who live in the building and make themselves available day and night. Although some courses focus on "pure" math or physics, most are in the problem-solving realm of what Hahne calls "relevant" mathematics—for example, related to bioinformatics, finance, or astronomy. When physicist Robert de Mello Koch of the University of the Witwatersrand in Johannesburg teaches electromagnetism, he avoids textbook tutorials, instead assigning challenging problems and projects such as building electroscopes out of soda cans. Understanding "the magic of AIMS" is only possible if you spend time with students, he says, "to see how hard they engage with the material, how far they manage to go, and how much it changes them."

Students appreciate their constant access to lecturers, a far cry from most of their university experiences. "The openness of lecturers, the nature of the material, the language—AIMS is completely different," says Khaleel.

This April, having completed their classes, the students were all working on final essays: Audry Ayivor of Ghana tackled topology in the library while Namayanja sipped tea as she explored a bioinformatics problem.

Even though AIMS is at a beach resort near Cape Town, this year's class has no South African students—and previous classes had only a handful of them—in part because the nation's talented math graduates are quickly hired by industries to fulfill diversity goals. In an effort to attract more South Africans, AIMS plans to start a separate Honors program in biological mathematics next year.

This month, Hahne says, AIMS is addressing another shortcoming by opening a new Research Center in two renovated buildings across the street. Visiting scientists, with joint appointments at AIMS and other universities, will try to create a synergy with the institute by enlisting students to help with research projects.

Although the goal of developing an Albert Einstein in Africa may not happen in their lifetimes (see main text), AIMS students don't discount the possibility. "There is a mountain of talent on this continent, but young people need opportunities to excel," says David Unuigbe of Nigeria, who wants to pursue solid-state physics at the University of Cape Town. And Ethiopian student Amasalework Ayele Ejigu believes AIMS will help students find such chances, ultimately benefiting the whole continent: "We are finding an African unity through mathematics."

–ROBERT KOENIG
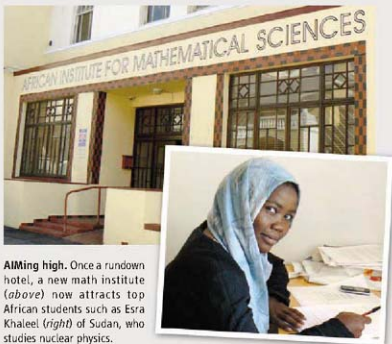
AIMing high. Once a rundown hotel, a new math institute (*above*) now attracts top African students such as Esra Khaleel (*right*) of Sudan, who studies nuclear physics.

Ministerial Council on Science and Technology, a group that included Turok began visiting math departments keen to become part of an African Mathematical Institutes Network (AMI-Net).

Turok is convinced that rather than adapting existing structures, it would be better to create something new in each place: a separate institute, within the university infrastructure but autonomous, to insulate it from any political pressure. "There's a better chance of success if it's managed as a franchise," he says. Not everyone agrees that starting from scratch makes financial sense. Although eager to replicate AIMS, Philippe Mawoko, a mathematician from the Democratic Republic of the Congo who works for the African Union's main development arm on science and technology initiatives, says it would be sensible "to use existing institutions as much as possible. Building completely new institutes would require more funding and legal work."

Since giving his TED talk in February, Turok says there has been "an amazing reaction." He says he's had discussions with Google, the Bill and Melinda Gates Foundation, the U.K. government, accounting firm PricewaterhouseCoopers, and Barclays Bank, as well as endorsements from the likes of Richard Branson, Bob Geldof, and Forest Whitaker. With $2.3 million worth of donations committed, he's aiming to raise $150 million over the next 5 years.

On 11 May, coinciding with the opening ceremony for a new research arm of AIMS, all parties involved in AMI-Net will gather in Cape Town to decide how to move forward. The first new institute, AIMS-Abuja, will open in July on the campus of the African University of Science and Technology, a World Bank project in Nigeria's capital. Turok says that once AMI-Net has enough money in place, it should quickly develop a model plan for how to set up new institutes and then embark on AIMS-3. "We have a huge responsibility to get it right. We have to make it work. It's the opportunity of a generation."

–DANIEL CLERY

With reporting by Robert Koenig.
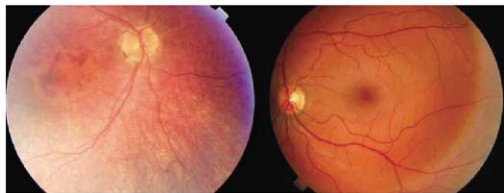
Dimmed sight. A retina affected by LCA2 (left) compared with a normal retina.

GENE THERAPY

Two Teams Report Progress in Reversing Loss of Sight

The first safety trials of gene therapy for a degenerative eye disease produced good results in adults; researchers now intend to treat children

At a vision research meeting this week in Florida, scientists made a stunning announcement: In two independent studies, gene therapy has partially restored the sight of four young adults born with severe blindness. After receiving a single injection of a solution containing a curative gene months ago, the patients can see more light. Some of them can now read several lines of an eye chart. And two who had previously stumbled through an obstacle course can now navigate through it.

The patients are still legally blind; they cannot read even a large-print book. Still, researchers are calling this first-ever test of gene therapy to treat a retinal disease a major advance. (The results were published online this week by the *New England Journal of Medicine*.) The patients' eyesight wasn't expected to improve much, if at all, because the disease was treated at an advanced stage. However, the two research teams—from the University of Pennsylvania (Penn) and Children's Hospital of Philadelphia (CHOP), and University College London—say the benefits may be much greater when they treat young children with this disorder, called Leber congenital amaurosis 2 (LCA2).

Gene therapy researchers hailed the news as a boost for their field. "It's a very promising development," says Arthur Nienhuis of St. Jude Children's Research Hospital in Memphis, Tennessee, president of the American Society for Gene Therapy. The studies were only designed to test safety. But if the hint of efficacy holds up in larger trials, the results could pave the way for using gene therapy for other inherited eye diseases as

well as for common ones such as macular degeneration, eye researchers say. "It is a marvelous thing for the field and for the future," says Paul Sieving, director of the U.S. National Eye Institute (NEI) in Bethesda, Maryland.

Not everyone who helped develop this treatment is sharing in the celebration. A third group that was part of a team that pioneered the work in dogs 7 years ago has been sidelined. Although the group's members at Penn and the University of Florida have results from an adult clinical trial, they have been forbidden by NEI, which funded the work, to discuss their findings. According to Sieving, the agency is waiting for an examination of safety and efficacy data to be completed.

Untreatable

The eye has long been an attractive target for gene therapy because the test gene is expected to be confined within the eye and because improved sight can be measured with precision. Interest grew in the 1990s, as dozens of genes for inherited retinal diseases were identified. They include LCA, a catchall term for severe blindness disorders affecting about 3000 Americans that lead to atrophy of the light-sensing photoreceptor cells in the retina. Children with LCA begin losing their sight at birth or soon thereafter; by age 40, they are completely blind. There is no treatment.

One form of LCA (LCA2) is caused by a defect in retinal pigment epithelium 65 (RPE65), a gene that is critical for the single

> *"It is a marvelous thing for the field and for the future."*
> —PAUL SIEVING, NATIONAL EYE INSTITUTE

layer of cells that line the back of the eye and nourish the photoreceptor cells. RPE65 codes for an enzyme that helps convert vitamin A into a form used to make rhodopsin, a pigment that absorbs light. Without rhodopsin, the photoreceptor cells eventually die. But this can take decades, so giving younger patients a working copy of RPE65 should restore photoreceptor function. The cells "just need to be resupplied [with pigment] and they're back in action again," says Michael Redmond of the NEI intramural program, whose lab discovered RPE65 and studied it in mice.

In 1998, gene therapists got a lucky break when researchers found that some Briard dogs carry the RPE65 mutation. A consortium formed to study gene therapy in these dogs. It included Cornell University veterinary researchers; scientists at the University of Florida, who developed the adeno-associated viral vector; and several Penn scientists. The Penn group included Samuel Jacobson, a clinician who runs a center for inherited retinal diseases, and two younger researchers: Albert Maguire, a retinal surgeon, and his wife, Jean Bennett, a molecular geneticist, who coordinated the studies. In 2001, the consortium reported in *Nature Medicine* that treating three young dogs with RPE65 mutations restored enough sight to allow them to avoid objects and walk through a maze. Their most famous subject, named Lancelot, became a poster dog, visiting members of the U.S. Congress to promote gene therapy.

In June 2005, Jacobson proposed a clinical trial for LCA2 in young adults to the National Institutes of Health's Recombinant DNA Advisory Committee (RAC), which reviews gene therapy trials. Bennett was not part of the proposed trial. According to Jacobson, she preferred to begin the trials in children. "There was a real difference in medical-ethical feeling," comments Jacobson, who received clearance to go ahead with the adult trial.

In the meantime, Katherine High, director of a new gene therapy center at CHOP, made Bennett and Maguire an offer: CHOP's center would fund a separate trial on children with Bennett and Maguire as leaders. High had raised money to start the center after a company sponsoring her own work on hemophilia dropped out. She decided gene therapy for LCA2 in children was a good trial to start

with. "The sooner you can intervene, the younger the patients, the better," High says.

In December 2005, Maguire proposed a pediatric trial to the RAC. He suggested that beginning with children made sense because it wasn't clear that adults would benefit. "Performing a study in older adults with devastated retinas may not provide adequate information to proceed with further studies in a pediatric population," Maguire said in an archived Webcast of the meeting. Jacobson, who was surprised by this presentation, notes that Maguire's appeal included drama: comments from a couple whose blind 1-year-old son had been seen by Maguire's team. The father wept at the microphone while his wife stood by with their baby. RAC approved the proposal.

The two groups—the Bennett-Maguire team, working through CHOP with private funding, and Jacobson's team with an NEI grant—moved ahead separately, although they briefly considered using a joint data and safety monitoring board (DSMB). NEI decided this wouldn't work, partly because the CHOP trial was moving faster, says NEI's Maryann Redford. The CHOP team and Italian collaborators began recruiting patients from the clinic of Francesca Simonelli at the Second University of Naples in Italy. Jacobson's NEI-funded team had a setback: It had to wait a year while the NEI approved a funding agreement for their trial.

Then early last year, a late entrant joined the race. Robin Ali's team at University College London, which had been working on gene therapy for other retinal diseases, wanted to launch an LCA2 trial because "we knew this would be a very good system to test efficacy," Ali says. They decided to treat adults first, Ali says, even though they knew results would be more dramatic with younger patients. It was the first time a human retina had been injected with a gene therapy vector, he notes. "If it had been a catastrophe," he says, "we would have been open to the criticism that we rushed into this with an 8-year-old."

The U.K. team treated its first patient in February 2007. The two U.S. teams did not begin until October. In the end, CHOP started

with adults, too—a 19-year-old and two 26-year-olds—because recruiting younger patients proved difficult and because adults can have more viable retinal tissue than they had thought, says Maguire. The injection caused no apparent immune response; one patient in the CHOP trial developed a pinhole-size macular hole, possibly from the injection surgery, but it didn't affect vision.

A few weeks later, a test based on pupil constriction showed that the patients could

Early intervenors. Penn's Jean Bennett and Albert Maguire argued that safety tests for LCA2 gene therapy should be done in children.

detect three times more light. Performance on eye charts was compelling, too, although such tests can be influenced by the placebo effect. Two patients who had been limited to detecting hand motions were able to read several lines of an eye chart. And a 26-year-old patient who initially could not navigate an obstacle course was able to make his way through.[*] "I couldn't believe it," Bennett says.

The London team, which used a weaker promoter sequence to drive gene expression,

saw improvement in only one patient, an 18-year-old who had the most intact retinal tissue, Ali says. He was no better at reading an eye chart, but his light perception improved 100-fold. In a video before treatment, he stumbles through a dimly lit, simulated night street scene, bumping into walls several times in 77 seconds. Six months after the injection, he breezes through in 14 seconds. "It's more than we could have hoped for," Ali says.[†]

Jacobson says the results from his NEI-funded study's first patient, treated in October, were also "very encouraging." His team drafted a case report in December, but NEI told them to wait until data for their next two patients—like the CHOP team's patients, treated in December and January 2008—had been reviewed in May by the study's DSMB. NEI is proceeding with caution, says Redford. The institute, however, issued a press notice last week describing the Penn trial and noting that NEI has spent $124 million on basic and clinical research on RPE65.

The other two teams say they sympathize. "It's very frustrating for everybody because all of us want to hear about their data," says Bennett. The competition between three groups has been healthy despite the disappointment for some, says eye disease geneticist Stephen Daiger of the University of Texas Health Science Center in Houston, who knows both U.S. groups. "It pushed everyone to work harder and faster in getting trials going. In the long run, that is great for all concerned."

The U.K. and CHOP teams are moving on to younger patients; Bennett and Maguire this month will treat an 8-year-old. They are also looking ahead to their next gene-therapy trial. Bennett says she and her Italian collaborators are thinking about recessive Stargardt disease, a disorder that affects about 27,000 Americans and involves a defect in a gene in photoreceptor cells.

Now that gene therapy for an eye disease has shown signs of helping the first set of patients, Bennett says, "it opens up so many opportunities." —JOCELYN KAISER

[*] content.nejm.org/cgi/content/full/NEJMoa0802315/DC1   [†] content.nejm.org/cgi/content/full/NEJMoa0802268/DC1

TB or not TB. That's the question for those tracing the origins of this mycobacterium.

Tuberculosis Jumped From Humans to Cows, Not Vice Versa

Humans have been cozy with their cows for almost 10,000 years—milking them, herding them, and even sleeping with them for warmth. Many researchers have thought that cows also gave our ancestors a less welcome gift: their first exposure to the mycobacteria that cause tuberculosis (TB), a disease that kills 2 million people a year.

At the meeting, a DNA study of 10 species of mycobacteria showed that early humans were infected with strains of *Mycobacterium tuberculosis*, which cause TB, long before they began herding cattle. That suggests that it was humans who transmitted the disease to bovids and other animals. "TB spread from humans to animals," perhaps when modern humans emerged from Africa to spread around the globe, reported graduate student Luz-Andrea Pfister of Arizona State University in Tempe in her talk.

Pfister compared the complete sequence of the same 60 genes in 10 species of mycobacteria from humans and animals. She sorted the bacteria into evolutionary trees and dated the splits among strains by assuming a constant mutation rate. She calculated that a proto *M. tuberculosis* split the mycobacterium that causes leprosy about 36 million years ago, perhaps in a primate ancestor of humans. She also estimates that *M. tuberculosis* in humans gave rise to *M. bovis*, the strain that infects cattle, about 113,000 years ago, and to the strains in several other mammals about 90,000 to 100,000 years ago. Those dates are based on an *Escherichia coli* mutation rate, because no rate has been calculated for mycobacteria. So

the precise dates may change, says Pfister, who is working to calculate a more precise mycobacterial clock.

Pfister's data suggest that humans gave the disease to cows; they also challenge a long-held view that *M. tuberculosis* descended from a single bug that flourished 20,000 to 35,000 years ago. That view was recently thrown into doubt by work by Cristina Gutierrez, an evolutionary mycobacteriologist at the

Institut Pasteur in Paris, and others who found unexpectedly diverse DNA sequences in African strains of *M. tuberculosis*. That raised the possibility that *M. tuberculosis* was more ancient, in order to have accumulated such diversity (*Science*NOW, 19 August 2005: sciencenow.sciencemag.org/cgi/content/full/2005/819/2). Pfister's findings, says Gutierrez, confirm "the emergence of variants of the *M. tuberculosis* complex long before the domestication of animals."

The genetic work also fits with recent fossil data showing putative TB lesions of great antiquity. For example, last year researchers reported such lesions in a 500,000-year-old archaic human fossil from Turkey, although without ancient DNA it's difficult to be sure TB was the culprit. Gutierrez thinks the main question now is "how could the first transmission from hominid to cow occur so long before their domestication?" TB transmission usually requires prolonged, close contact—but not always, says evolutionary osteopathologist Bruce Rothschild of the University of Kansas, Lawrence. He points out that "primates in zoos can get TB from humans who spit on them." That raises the specter of a wheezing human ancestor who spat out an insult—in the form of TB—on some hapless bovid long ago.

Australopithecus Not Much of a Nutcracker

In 1959, famous South African paleoanthropologist Phillip Tobias dubbed a newly discovered skull from Olduvai Gorge in Tanzania "Nutcracker Man," because of its huge cheek teeth, thick tooth enamel, and massive jaw. Ever since, researchers have thought that the teeth and jaws of these robust australopithecines were specialized to chomp on hard nuts and seeds, whereas their more fine-boned cousins, the gracile australopithecines who eventually gave rise to humans, were generalists who ate fibrous fruit and plants.

Now, several researchers reported at the meeting, different analytical methods suggest that the diet of robust australopithecines wasn't so hard after all, and that robust and gracile hominids ate similar fare. Even Nutcracker Man's species, now called *Paranthropus boisei*, wasn't crunching nuts and small hard objects routinely, according to a new analysis of its tooth wear. "It looks more like it was eating Jell-O," says paleoanthropologist Peter Ungar of the University of Arkansas, Fayetteville.

These results "show clearly that simplistic ideas about australopithecine diets need to be revised," says paleoanthropologist David Strait of the University at Albany in New York.

Ungar and his colleagues exposed the soft side of Nutcracker Man's diet by scanning the molars of seven individuals of *P. boisei* from East Africa with a confocal microscope and analyzing microscopic wear and tear. The teeth lacked the telltale pits that come from eating small, hard nuts and seeds, as seen in gray-cheeked mangabeys and brown capuchin monkeys, the team reported in *PLoS One* this week. In fact, the robust australopithecine's pattern of wear resembles that of the gracile hominids, including *A. afarensis* from East Africa. At the meeting, Ungar reported in a talk that no australopithecines show signs of eating hard objects routinely, although *A. robustus* from South Africa may have done so seasonally.

A separate line of evidence from isotopic studies is also blurring the distinction between the diets of robust and

CREDIT: GETTY IMAGES/SCIANCECHICAL.COM

Snapshots From the Meeting >>

Evolution of gliding. In the tropical forests of Southeast Asia, more animals glide from tree to tree than anywhere else in the world: leaping lizards and geckos, 45 species of frogs, and snakes that can glide 100 meters. Researchers have proposed that more gliding species evolved there because Asia's taller trees give animals enough time to gain lift after they fall, and there are few vines to use to move between trees. At the meetings, a researcher who mapped the canopies of 11 forests around the world with laser equipment offered another explanation: rugosity, or vertical gaps between treetops of different heights.

Biological anthropologist Nathaniel Dominy of the University of California (UC), Santa Cruz, reported that Asian forests showed the most extreme vertical drops from treetop to treetop, especially in high canopies of forests where trees grow more than 50 meters tall. Falling can be fatal, Dominy says, so natural selection would strongly favor animals with the anatomy to recover by gliding. As a result, says UC Berkeley physiologist Robert Dudley, Asia has many gliders whereas African and American forests have only a few. The steep plunges may even have shaped Asian apes: Dominy proposed in his talk that gibbons have developed their unique mode of ricochet brachiation (rapid, arm-over-arm swinging) to navigate the vertical drops. "There's lots of monkeying around in tall forests," Dudley says.   —A.G.

Side by side. One of the hottest debates in paleoanthropology concerns the petite *Homo habilis*. Did it give rise to the long-legged human ancestor *H. erectus* about 1.8 million years ago in Africa? Or were the two species contemporaries who arose from an even earlier ancestor? At the meetings, paleoanthropologist Christopher Ruff of Johns Hopkins University in Baltimore, Maryland, offered the first analysis of the interior structure of arm and leg bones from *H. habilis*—and concluded that it and *H. erectus* moved in different circles.

By examining how bone density varies in cross sections of upper arm and thigh bones, researchers can see how mechanical loads are distributed on a hominin's limbs as it walks or climbs trees, for example. Ruff studied images of the cross sections from the arm and leg bones of *H. habilis, H. erectus,* chimpanzees, and 100 modern humans. His analysis is "convincing" that many features of *H. habilis* limbs are more primitive than those of *H. erectus,* says paleoanthropologist Henry McHenry of the University of California, Davis. Ruff suggests that *H. habilis* spent more time in the trees, supporting fossil evidence that the

Flying leap. Many animals, such as this sugar glider, evolved to navigate vertical gaps among the tall treetops of Southeast Asia.

two species lived in different habitats. And that, in Ruff's view, makes it unlikely that *H. habilis* gave rise to *H. erectus*.   —A.G.

Neandertal voice-over. When Neandertals talked, what did they sound like? Robert McCarthy of Florida Atlantic University in Boca Raton played a brief synthesized version of a Neandertal voice, based on his reconstruction of their vocal tracts. The vowel sound of "e" in "see" was less precise than ours. It also sounds higher-pitched to our ears, although that's an artifact of this particular sound, McCarthy says.

McCarthy and Philip Lieberman of Brown University reconstructed the shape of the Neandertal vocal tract from fossils of the face and neck bones and inferences based on modern humans. The frequencies that emerged most clearly from their throats were different from ours. As a result, Neandertals would not have been able to form the precise, rapid-fire phonemes of human speech, the researchers found.

Perhaps the biggest surprise: A 100,000-year-old early modern human from Israel, known as Skhul 5, had the same vocal traits as Neandertals. "Full-blown modern speech came relatively late," says McCarthy, perhaps between 100,000 and 40,000 years ago   —ELIZABETH CULOTTA

gracile australopithecines. Various plants absorb atmospheric carbon dioxide differently, and so by measuring the ratio of carbon isotopes in teeth, researchers can detect whether ancient hominins ate tropical grasses and sedges rich in carbon-13, or woody fruits, shrubs, and herbs with less carbon-13.

The isotopic data from the gracile australopithecines matches that of the robusts and suggests that both groups ate more diverse diets than expected, said paleoanthropologist Matt Sponheimer of the University of Colorado, Boulder, in his talk. His team analyzed the carbon isotopes in teeth from 20 robust and 25 gracile australopithecines from South

Africa. Both groups clustered in between animals that browse on fruit and leafy plants, such as fossil giraffes and fruit-eating antelopes, and animals that graze on fibrous

False teeth? Nutcracker Man's huge molars may not have been used to regularly crush hard foods.

grasses, such as fossil wildebeest and zebra. Sponheimer concluded that the robust australopithecines' cuisine included fruits from trees, bushes, and shrubs when they could get them but also less desirable grasses and sedges, similar to the diets of some baboons.

So why did robust australopithecines evolve such massive jaws and molars? The best guess at the meetings was that the extreme jaw-breaker anatomy may have been an adaptation to eat less desirable "fallback" foods in times of drought or seasonal stress. Gorillas, for example, have evolved sharp teeth to eat fibrous leaves when needed for survival, although they prefer and consume far more fruit and soft leaves. "The anatomy shows not what an animal eats but what it has the potential to eat," says Ungar.   —ANN GIBBONS

LETTERS

edited by Jennifer Sills

Ensuring Food Security

IN A RECENT PERSPECTIVE, "FOOD SECURITY UNDER CLIMATE CHANGE" (1 February, p. 580), M. E. Brown and C. C. Funk conclude that improved seed, fertilizer, land use, and governance lead to food security. I find these claims highly questionable. The green revolution model (monocultures of improved crops supported through high levels of agrochemical and other inputs) has done much to increase agricultural productivity. It does little, if anything, to increase food security.

Farmers in developed countries raise and sell crops but buy their food in supermarkets. Despite improved seeds and fertilizers, crops sometimes fail. When this happens, Western farmers receive government or other insurance payments. This scenario does not always apply in less developed regions or to subsistence farmers.

Across the wide scope of agriculture, there are plenty of ecologically sound, food-ensuring mechanisms. At the farm level, land modifications, climatically stable agroecosystems, plot landscape positioning, alternative crops or varieties, in-soil vegetative material, and well-placed biodiversity can all play a role in countering unfavorable climatic events.

Organized traditional societies avoid recurrent periods of starvation through multiple and overlapping mechanisms. For example, the Incas used crop varieties, communal irrigation, stone terraces, and plot scattering, along with community food storehouses, to lessen or mitigate famines.

Less organized farmers often rely upon a single mechanism. In wetter regions of West Africa, farmers plant rainfall-demanding rice along with drought-resistant cassava. Early European and traditional African societies placed a greater reliance upon climatically resilient tree crops. In early Europe, the acorn was the bailout crop (1). In Africa, there are a number of fruiting tree species that yield even when staple crops are lost (2).

Another advantage of alternative agricultural models—always in place, always functioning—is that they do not require monitoring or prediction.

PAUL WOJTKOWSKI

65 Dexter Avenue, Pittsfield, MA 01201, USA.

References
1. D. A. Bainbridge, *Ambio* **14**, 148 (1985).
2. F. A. M. Harris, S. Mohammed, *Ambio* **32**, 24 (2003).

Incan terraces. Societies throughout history have used a variety of methods to ensure food security. Stone terraces built by the Incas are one example.

Response

WOJTKOWSKI SEEMS TO ASSERT THAT MONItoring and prediction of variations in agricultural productivity are not only unnecessary, but actually a waste of resources for ensuring food security. He also seems to assert that we propose a one-size-fits-all, Western, agribusiness solution. We disagree with both of these assertions. Our suggestions regarding how to ensure adequate food availability in regions chronically food-insecure certainly include many of the food-ensuring approaches mentioned by Wojtkowski, such as land modifications and plot landscape positioning. Our piece also focused on other issues that affect the ability of the poorest and most vulnerable to access food, such as technological transformation. The sources of food insecurity are complex and will require complex solutions (1).

While there is a role for conventional seed, fertilizer, and other technology, many natural resource management techniques mentioned by Wojtkowski have been successfully implemented in the Sahel and can lead to incremental gains that help close the food security gap. Farmers in Africa have always had diverse agroecosystems, and examples abound of those who have increased their agricultural diversity and productivity. An excellent example is Niger, where tree planting and conservation have transformed highly degraded landscapes into productive agroecosystems. Farmers are able to produce more food on less land, more reliably (2, 3). Nevertheless, Niger faces chronic and mounting food-production challenges that will be difficult to meet through improved landscape ecology alone. In Niger, 20 years ago, fertility rates were seven children per

woman. Today, fertility rates are 7.2 per woman, and Niger has cultivated 91% of its potentially cultivatable land. Under current population and agricultural expansion rates, Niger will run out of new land to cultivate by 2015 (4).

Issues of food security are compounded by the impacts of poor governance on food access and utilization. Governance includes the education of women and children, provision of clean water and health care, and a stable functioning market system. Again, Niger is a good example of the complexity of the food security problem (4). There are still incredibly high levels of malnutrition in the country, particularly among children. In the 2004–2005 crisis, chronic malnutrition caused by poor child-feeding habits and an insufficiently diverse diet was exacerbated by declines in food production outside of the

country and high food prices. These changes in food availability and access caused massive increases in enrollments in child-feeding programs and a large increase in the need for humanitarian assistance.

Ensuring food security for all in the face of climate-caused reductions will require adequate food production through improved seed and fertilizer; better land use policies and good governance; as well as appropriate interventions, safety nets, and investments during crises. Africa's culture, landscape, and challenges are complex, and complex solutions integrating responses from the social, political, economic, and biophysical domains will be required.

MOLLY E. BROWN,[1] CHRISTOPHER C. FUNK,[2] JAMES VERDIN,[3] GARY EILERTS[4]

[1]SSAI/NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [2]Department of Geography, University of California, Santa Barbara, CA 93106, USA. [3]U.S. Geological Survey, Sioux Falls, SD 57198, USA. [4]U.S. Agency for International Development, Washington, DC 20523, USA.

References

1. OXFAM, Causing Hunger: An Overview of the Food Crisis in Africa, OXFAM briefing paper, July 2006; www.oxfam.org/en/policy/briefingpapers/bp91_africa_food_crisis.
2. R. Harris, "Niger's trees may be insurance against drought," National Public Radio, All Things Considered, July 2007; www.npr.org/templates/story/story.php?storyId=11608960.
3. The World Bank, "Agriculture for development" (World Development Report, 2008); http://go.worldbank.org/ZJJAOSUFU0.
4. G. Wentling, "Niger—Annual Food Security Report: Current Situation and Future Prospects" (USAID/WA Country Program Manager for Niger American Embassy, Niamey, Niger, 2008).

Coarse-Resolution Models Only Partly Cloudy

IN THEIR REPORT, "A MADDEN-JULIAN Oscillation event realistically simulated by a global cloud-resolving model" (14 December 2007, p. 1763), H. Miura et al. promote the impression that coarse-resolution climate models cannot simulate the Madden-Julian Oscillation (MJO). We would like to point out that some coarse-resolution climate models, using conventional parameterizations of convection and clouds, can represent the boreal winter and summer MJO with fidelity (1–9).

In several studies, one coarse-resolution atmospheric model validated the MJO in terms of convection, eddy stream function, and surface evaporation, and it was hypothesized that lack of air-sea interaction contributed to shortcomings in the MJO simulation (1). This hypothesis was later borne out, resulting in a more realistic MJO simulation (2). Subsequently, the model was used for idealized predictability studies that demonstrated

the potential for the MJO to be forecast with lead times of 15 to 30 days (3). Using a different set of models, for which more complete model diagnostics were available, important aspects of the MJO were realistically represented, including the relationships between convection and low-level moisture convergence, surface fluxes, the vertical structure of winds and divergence, and air-sea interactions (4). None of these relationships, including the spontaneous generation of MJOs, have been adequately demonstrated in Miura et al. to justify their claim of a realistic MJO or their inference that high-resolution models may be necessary to represent the MJO.

Other coarse-resolution simulations capture the northward propagating component of boreal summer intraseasonal variability (5–7). During both summer and winter, a realistic representation of the time-mean climate state is required to produce a realistic MJO (4, 8, 9). These works provide evidence that coarse-resolution climate models have been successful in understanding mechanisms involved in the propagation of the

MJO, and for exploring important applications of MJO variability and predictability.

KENNETH R. SPERBER,[1]* JULIA M. SLINGO,[2] DUANE E. WALISER,[3] PETER M. INNESS[2]

[1]Lawrence Livermore National Laboratory, Program for Climate Model Diagnosis and Intercomparison, Post Office Box 808, L-103, Livermore, CA 94551, USA. [2]National Center for Atmospheric Science and Walker Institute, University of Reading, Earley Gate, Reading RG6 6BB, U.K. [3]Jet Propulsion Laboratory, MS 183-505, California Institute of Technology, Pasadena, CA 91109, USA.

*To whom correspondence should be addressed. E-mail: sperber1@llnl.gov

References

1. K. R. Sperber, J. M. Slingo, P. M. Inness, K.-M. Lau, Clim. Dyn. 13, 769 (1997).
2. D. E. Waliser, K.-M. Lau, J.-H. Kim, J. Atmos. Sci. 56, 333 (1999).
3. D. E. Waliser, K.-M. Lau, W. Stern, C. Jones, Bull. Am. Meteorol. Soc. 84, 33 (2003).
4. K. R. Sperber, S. Gualdi, S. Legutke, V. Gayler, Clim. Dyn. 25, 117 (2005).
5. X. Fu, B. Wang, J. Clim. 17, 1263 (2004).
6. K. Rajendran, A. Kitoh, J. Clim. 19, 366 (2006).
7. K. R. Sperber, H. Annamalai, Clim. Dyn., 10.1007/s00382-008-0367-9 (2008).
8. P. M. Inness, J. M. Slingo, J. Clim. 16, 345 (2003).
9. P. M. Inness, J. M. Slingo, E. Guilyardi, J. Cole, J. Clim. 16, 365 (2003).

CORRECTIONS AND CLARIFICATIONS

Newsmakers: "Thermometer kings" (4 April, p. 29). Richard Porter's thermometer collection did not include an earring thermometer from a 1650 whaling ship.

News of the Week: "China's modern medical minister" by R. Stone (28 March, p. 1748). Chinese minister Wan Gang's name was misspelled.

Random Samples: "Homeward buzz" (March 14, p. 1465). The article referred to a "colony of 750,000" bees. The collection is actually 56 colonies of 14,000 bees each. Each colony has its own queen.

Reports: "Asphericity in supernova explosions from late-time spectroscopy" by K. Maeda et al. (29 February, p. 1220). In author affiliation 16, the name of the institution should have been Institute of Space and Astronautical Science (not Astronomical Science). In ref. 8, the order of author names was incorrect. The reference should read as follows: 8. R. Buras, H.-Th. Janka, M. Rampp, K. Kifonidis, Astron. Astrophys. 457, 281 (2006).

Reports: "Nonadiabatic interactions in the Cl + H2 reaction probed by ClH2− and ClD2− photoelectron imaging" by E. Garand et al. (4 January, p. 72). This research was supported by Air Force Office of Scientific Research grant F49620 03 1 0085 (D.M.N.), NSF grant CHE0413743 (M.H.A.), and Office of Naval Research grant N000140510460 (D.E.M.). E.G. thanks the Natural Science and Engineering Research Council of Canada for a postgraduate scholarship.

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Long-Lived Giant Number Fluctuations in a Swarming Granular Nematic"

I. S. Aranson, A. Snezhko, J. S. Olafsen, J. S. Urbach

Narayan et al. (Reports, 6 July 2007, p. 105) reported giant number fluctuations attributed to curvature-driven active currents specific for nonequilibrium nematic systems. We present data demonstrating that similar results can be found in systems of spherical particles due either to inelastic clustering or persistent density inhomogeneity, suggesting two alternative explanations for their results.

Full text at www.sciencemag.org/cgi/content/full/320/5876/612c

RESPONSE TO COMMENT ON "Long-Lived Giant Number Fluctuations in a Swarming Granular Nematic"

V. Narayan, S. Ramaswamy, N. Menon

On the basis of experiments using monolayers of spherical grains, Aranson et al. suggest that the giant number fluctuations we observed in active granular rods may be explained by static inhomogeneity or inelastic clustering. We refute these alternative explanations and underline the evidence that the fluctuations originate in nematic ordering.

Full text at www.sciencemag.org/cgi/content/full/320/5876/612d

Response

WE DEMONSTRATED THAT A GLOBAL CLOUD-resolving model (GCRM) can simulate the realistic evolution of a single Madden-Julian Oscillation (MJO) event, including its internal structures. We did not claim that GCRMs provide a full solution to the MJO problem or that conventional general circulation models (GCMs) cannot simulate the MJO.

The essential mechanisms of the MJO are not yet comprehensively understood, and consequently, whether GCMs and GCRMs can fully simulate the MJO remains undetermined. The MJO simulations in our Report demonstrate that a GCRM can

marginally reproduce the internal cloud systems and overall structure of a single MJO event over a period of one month (1). Finer internal structures and topographic effects were also better simulated because of the finer horizontal resolution. Contrary to the assertion of Sperber et al., however, we did not claim that either GCRMs or GCMs embedding two-dimensional cloud-resolving models were the only way to produce realistic MJO simulations.

Most of the papers cited by Sperber et al. stated that air-sea interaction is necessary to simulate the MJO by GCMs. However, we believe that this point is still controversial (2, 3). We have studied this issue and performed a sensitivity test with a 14-km horizontal grid. The time evolution of an MJO event did not differ very much between the simulations with the observational time varying sea surface temperature (SST) and the fixed SST, though the convective activity of the MJO was weaker in the fixed SST run. A GCM embedding a two-dimensional cloud-resolving model also simulated the MJO without feedback from intraseasonal perturbations in SST (4). These results sup-

port the hypothesis that the MJO is inherently an atmospheric mode, even if it can be modified and perhaps amplified by air-sea interactions (5, 6).

HIROAKI MIURA,[1,2*] MASAKI SATOH,[1,3] TOMOE NASUNO,[1] AKIRA T. NODA,[1] KAZUYOSHI OOUCHI[1]

[1]Frontier Research Center for Global Change, Japan Agency for Marine-Earth Science and Technology, 3173-25 Showamachi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan. [2]Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523–1371, USA. [3]Center for Climate System Research, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8568, Japan.

*To whom correspondence should be addressed. E-mail: miura@atmos.colostate.edu

References
1. D. L. Hartmann, H. H. Hendon, Science 318, 1731 (2007).
2. Climate Change 2007, "The physical science basis," Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter 8, Climate Models and Their Evaluation, 8.4.8 Madden-Julian Oscillation (p. 625).
3. C. Zhang, Rev. Geophys. 43, RG2003 (2005).
4. M. F. Khairoutdinov, D. A. Randall, C. DeMotte, J. Atmos. Sci. 62, 2136 (2005).
5. D. E. Waliser, K.-M. Lau, J.-H. Kim, J. Atmos. Sci. 56, 333 (1999).
6. P. M. Inness, J. M. Slingo, J. Clim. 16, 345 (2003).

NEUROSCIENCE

On Deciding How to Do unto Others

Prashanth Ak

Naturalized ethics, the idea that principles of natural science bear on normative ethics, faces two long-standing objections. The naturalistic fallacy cautions that good, in the moral sense, cannot be defined from natural properties. Hume warned against deriving an ought (as in how people ought to act) from an is (how people actually act, for instance). Most of those who seek to naturalize ethics are familiar with these arguments but maintain that scientific findings can have a profound impact on our understanding of morality and ethics.

The question of whether ethical concepts have innate bases or are acquired has echoes of the nature-versus-nurture question and carries much the same ideological baggage. Whether it is our natures or our cultures that make us who we are has been central to all sorts of intense debates, on topics including the ideal political system, effective means of teaching, and crime and punishment. Discussions of the nature of morality exhibit similar polarity, with some boasting of their indifference to neuroscience and others embracing it wholeheartedly.

Such issues are not the explicit focus of Donald Pfaff's *The Neuroscience of Fair Play* but nevertheless surround the book. Pfaff, a neurobiologist at the Rockefeller University, reasons that because the golden rule—treating others as one would like to be treated, stated variously—is universal, it could have neurobiological bases.

Pfaff is careful to clarify that he does not think there is a "signaling circuitry devoted to ethics." What he aims to do is present possible neural and genetic mechanisms underlying the golden rule. He also clarifies that he is not hoping to settle disputes among evolutionary biologists, psychologists, and philosophers or address issues in real-life, complex social processes.

The book starts off with an arresting example, the heroism of Wesley Autrey, who threw himself on top of a stranger in the New

> **The Neuroscience of Fair Play**
> Why We (Usually) Follow the Golden Rule
> **by Donald W. Pfaff**
> Dana Press, New York, 2007. 248 pp. $20.95.
> ISBN 9781932594270.

York City subway to save the person from being crushed by an oncoming train. Pfaff hypothesizes that such altruism is due to brain mechanisms that override self-preference and blur the boundaries between the self and the other through a "loss of social information." He conjectures that it depends in part on neurobiological mechanisms for fear, supplemented by neurohormonal bases of sexual and parental behaviors, and that departures from altruistic behavior are due to similar neurogenetic bases of antisocial behaviors. Pfaff suggests that the capacity of a person to behave according to the golden rule depends on a balance—properly, an imbalance—among social behavioral mechanisms in which those producing prosocial actions outweigh those producing antisocial actions.

The author presents the science clearly and in sufficient detail to enable readers of

George Bernard Shaw's advice: "Do not do unto others as you would that they should do unto you. Their tastes may not be the same."

The reviewer is at the Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA. E-mail: prashanth@ias.edu

Science to make up their own minds about the plausibility of this proposal. The book doesn't offer a comparable level of technical detail when it comes to the relationship of biology to behavior—and there is even less on the connections between behavior and ethics. Nevertheless, readers won't have any difficulty evaluating the validity of the hypothesis presented, because Pfaff spells his ideas out clearly enough.

The presentation contains a few factual errors that might bother the aficionados. For example, the Buddhist concept of anatman does not mean "universal spirit" but "no-soul." (Buddhism explicitly negates ideas of universal spirit.) The differential expression of the vasopression receptor gene is not due to sequence changes in the promoter, rather the changes lie in the microsatellite region about 500 base pairs upstream of the gene. But such errors are minor in the context of the main points of Pfaff's speculations and do not contradict them.

Pfaff's broad-brush treatment of altruism, however, is bound to bother quite a few readers. He states that the golden rule is, "in biological terms, 'reciprocal altruism' " (costly acts that benefit others, performed with the expectation of reciprocation). Neither type of altruism mentioned in the book, kin or reciprocal altruism, explains human generosity to non-kin when reciprocation and repeated interaction are unlikely and costs are high—such as in the example of Wesley Autrey.

In repeated situations involving reciprocal altruism (such as the iterated prisoner's dilemma game), a common optimal strategy is tit for tat: begin by playing nice and then do what the other player does, nice or not. This is not quite the golden rule but rather "a tooth for a tooth" ethic. Other strategies that are less susceptible to errors exist, such as "win-stay, lose-shift": repeat previous move whenever doing well; if not, change. These too are not easy to formulate as the golden rule.

Further, in such scenarios, especially with multiple players, the presence of a single selfish player is enough to unravel cooperation. Because the golden rule (in its various formulations) has persisted over two millennia, chances are that something more stable than reciprocal altruism, with its easily frayed cooperation, underlies it. Reciprocal altruism can be made more stable by adding features such as cooperation contingent on the recipient's reputation for cooperation (indirect reciprocity) or penalties for not

cooperating (strong reciprocity). If reciprocal altruism as Pfaff uses it is meant to include these forms as well, considerably more sophisticated cognitive mechanisms than those outlined in the book are required.

The book's concluding pages, which present possibilities of neuroscientific interventions to favor prosocial behaviors and reduce violence, will raise a few eyebrows if not hackles. However, that's a likely consequence of books that bring topics forward for debate and, hopefully, further our understanding. *The Neuroscience of Fair Play* successfully highlights important issues in a young field of inquiry. Although readers may find much to disagree with in Pfaff's account, clear formulations of their objections will help advance the study of possible biological bases of morals.

10.1126/science.1157089

COGNITIVE SCIENCE

Rethinking Folk Psychology

Erik Myin

The field of research on how we mutually understand each other has been one of the most active in the cognitive sciences. Even the general public has become aware of some of its findings, such as the problems people with autism have in interpreting actions and the existence of "mirror neurons"—neurons that fire both when you perform an action and when you see someone else performing that action. The research domain has been organized around the central idea that people's capacity for mutual understanding is structurally similar to scientific understanding. We see the movements that our fellow human beings make not as mere motions but as actions driven by desires and informed by beliefs. We explain and predict each other's behavior, so common wisdom goes, in terms of a folk psychology of beliefs and desires—just like physicists explain and predict visible phenomena such as droplets in a cloud chamber by reference to invisible electrons.

Traditional explanations of our folk psy-

chological capacities split on whether the crucial mechanism for understanding others is a result of genuinely theorizing about their beliefs and desires (a theory of mind) or of simulating these. Nevertheless, nearly all researchers in the tradition invoke complex "mindreading" machinery, operating behind the scenes. Moreover, it is generally assumed that this cognitive machinery has a strong innate component. The machinery must have been present in our evolutionary precursors, so a common argument goes, or else some of their well-established capacities— e.g., deception, social learning of tool use, social cooperation, the emergence of symbolic language—cannot be accounted for.

In *Folk Psychological Narratives*, Dan Hutto presents an alternative conception of folk psychology as well as a thorough critique of its traditional treatment in the cognitive sciences. Hutto, a philosopher of psychology and professor at the University of Hertfordshire, rejects the idea that our stance toward each other is genuinely "theoretical." Mutual understanding, he contends, typically is obtained through "second-person encounters," face-to-face situations in which people ask about and provide reasons for why they acted as they did. Most often, we have a situational understanding of people's behavior: we understand their actions as being in accordance with the contextual demands of a more-or-less standard situation. Only when someone's behavior deviates from our expectations about what would befit the context does the need for an explanation in terms of special reasons ever arise.

According to Hutto, mastering folk psychology (which he considers as the art of providing and asking for reasons) is a late achievement. Its final acquisition is realized through encounters, in conversations or in the context of stories, with narratives about people who act for reasons, "folk psychological narratives." The book describes the various steps along the gradual development (through a series of transformations of simpler capacities) of both the capacity to act for a reason and the capacity to understand acting for a reason.

In sketching this cognitive trajectory, Hutto insists on the importance of a public language as the means of providing both the necessary content and form for the kind of

Mutual miming. This is one of our capacities that Hutto uses to account for the emergence of human technical, social, and linguistic skills.

complex thinking that having and understanding reasons require. Language-less cognitive abilities exist, fostered by abilities of recreative imagination (reenacting previous perceptual experiences), but they fall far short of the complexity required to read minds in terms of beliefs and desires. As Hutto says in a characteristic formulation, "This is devastatingly bad news for those inclined to believe in the existence of nativist mindreading mechanisms of any sophistication."

Hutto estimates that the discursive practices that were themselves only preconditions for folk psychology arrived on the scene "long *after* the last possible date for universal anatomical change in early humans." This timing offers an empirical reason for rejecting innate mindreading in explanations of deception in hominids, social learning of tool production, social cooperation, and the emergence of symbolic language. Hutto provides sketches of alternative explanations for these phenomena in terms of simpler cognitive abilities.

Although *Folk Psychological Narratives* is engagingly written, Hutto's account will probably be easier for those readers who are to some degree already familiar with the fundamental debates that have raged in cognitive science through recent decades.

Through the presentation of his alternative conception of folk psychology, tracing out its development, and the detailed elaboration of its explanatory virtues, the author succeeds at striking a balance between the constructive and the critical. His challenge to the theory of mind tradition is formidable. At the same time, Hutto's "narrative practice hypothesis"—that the normal route by which children acquire folk psychological abilities is through their encounters with stories about people who act for reasons—offers the field a promising basis from which to reorient itself.

10.1126/science.1157120

ECONOMICS

Linking Natural Resources to Slow Growth and More Conflict

Natural resources do not necessarily spell doom for development.

C. N. Brunnschweiler[1]* and E. H. Bulte[2]

The appreciation for natural resources as a driver of economic development has undergone a dramatic change in the past decades. Although an abundance of resources was generally perceived as advantageous until the 1980s, an influential literature emerged in the 1990s that reached seemingly opposite conclusions. The phrase "natural resource curse" was coined, and perhaps because of its paradoxical connotation, caught on in both academic and policy circles. Two prominent "dimensions" of the resource curse include the association of resources with slow economic growth [a literature inspired by Sachs and Warner (1)] and with armed civil conflict [a literature mainly inspired by Collier and Hoeffler (2)].

The causal mechanisms linking resources to slow growth and more conflict are ill understood. It is often argued that resource-rich economies suffer from weak leadership, rent seeking, and failing institutions (3). This may be either because resource profits (rents) trigger "rentier state" dynamics and the associated disconnect between the rulers and the people, or because resource rents enable autocratic and unaccountable rulers to oppress opposition (4). Resources may also invite conflict if greedy rebels seek profitable looting opportunities. Finally, dependency theories of development predict that the strategic and commercial value of resources may affect politics

and economic outcomes in developing countries, as they are of interest to powerful nations and corporations. The stories associated with the curse are compelling, and

Natural resource dependence, abundance, and economic growth. Regression fits of natural resources and economic growth 1970–2000. (**Top**) Natural resource dependence in 1970; (**bottom**) World Bank total natural wealth data (log values) measured in USD per capita in 1994.

ample anecdotal evidence exists to lend credibility to the key ideas.

The curse is now an immensely popular research topic and receives serious attention from multilateral agencies and nongovernmental organizations (NGOs) (5). Its increasing status within the development community is reflected by the fact that international organizations are providing advice to resource-rich developing countries on how to manage their resource base (reducing reliance on the primary products sector) and revenue streams to exorcise the curse. Some of the proposals are quite radical, such as the

one suggesting to first distribute resource profits to the people and then to tax them back (6). Increasingly, there are calls to regulate international trade to face certain manifestations of the curse head-on (e.g., the Kimberlite initiative dealing with "blood diamonds").

But how robust is the evidence for the curse? We believe it is actually weaker than generally perceived. A key problem of most existing analyses is that the common resource variable used in cross-country regression models is endogenously determined, and itself not invariant with respect to changes in institutional quality or conflict (the variables it is supposed to adversely affect). If so, existing empirical results would be biased.

The standard resource variable used by Sachs and Warner, as well as by Collier and Hoeffler, is primary exports divided by a measure of national income. It thus captures the resource dependence of economies, rather than abundance. A negative correlation between this variable and growth could mean that resources lead to slower economic growth, as suggested by the curse proponents. Alternatively, it could mean that poor economic development policies—leading an economy to become dependent on its primary exports—dampen growth. Similarly, although a negative correlation between the resource variable and institutional quality may imply that resources undermine institutions, it might also capture that the resource sector is the "default sector" in the absence of decent institutions when nobody is willing to invest in alternative forms of capital. Finally, a positive correlation between the resource variable and conflict may indeed mean that

[1]CER-ETH (Center of Economic Research at the Swiss Federal Institute of Technology), Zurich, Switzerland. [2]Development Economics Group, Wageningen University, Netherlands, and research fellow of the Oxford Center for the Study of Resource-Rich Economies (Oxcarre), Oxford, UK.

*Author for correspondence. E-mail: cbrunnschweiler@ethz.ch.

resources trigger conflict. But it may also be the case that conflict makes countries dependent on resource extraction—the default activity that still takes place after other economic sectors (more mobile or, perhaps, better linked to the rest of the economy) have come to a stop. If so, resources are not a curse to development, but rather a safety net to support people and economies even under adverse circumstances. The nature of the causality is, therefore, a concern.

The importance of finding an appropriate proxy for resource endowments, as well as the consequences of this proxy for econometric results, is illustrated by the simple example in the figure on page 616. At the top, a regression fit of the conventional resource variable—primary exports divided by GDP at the start of the period—on economic growth between 1970 and 2000 results in the usual negative "curse" relation. At the bottom, however, a new resource wealth variable is used, and the result is reversed, showing a positive correlation between resource abundance and growth—the curse disappears!

The resource variable used in the bottom figure is one of several made available by the World Bank [7]. They capture the discounted value of expected resource rents for a future period of 20 to 25 years, calculated in U.S. dollars (USD) per capita for 1994. Contrary to the standard resource variable (which captures flows), these wealth variables estimate resource stocks—both aggregate and divided by type, such as mineral or cropland assets. They therefore offer more intuitive variables to measure abundance.

In more extensive tests, we used standard econometric techniques to shed light on the causation issue. We used a so-called instrumental variable approach to isolate effects of income and resource dependence on conflict, rather than the reverse effect [8]. We do the same to isolate the effect of dependence and institutions on economic growth. We also consider the effect of resource abundance on growth and conflict, using the World Bank resource variables. A summary of representative results, including technical details about the estimation approach, is available on Science online [9]. Our main results are as follows. If we adopt the conventional methodology—that is, simply assume that resource dependence is an exogenous explanatory variable in growth and conflict regressions—then our data reproduce the conventional curse results. In other words, there appears to be a significant negative relation between resources and growth, and a positive rela-

tion between resources and the probability of conflict. However, inspection of these results suggests that the conventional methodology is flawed and can produce biased results. Specifically, as discussed above, resource dependence is endogenous in the regressions [9].

After addressing the endogeneity problem, the correlation between resource dependence on the one hand, and conflict and slow growth on the other, vanishes. The correlation between resource dependence and slow growth and conflict, therefore, does not imply causation from the former to the latter. Instead, causality appears to be running from weak institutions and conflict to resource extraction as the default sector, which produces resource dependence as the final outcome. Resource dependence appears as a symptom, rather than a cause of underdevelopment. These results are robust to alternative model specifications [9].

However, as already suggested by the simple results at the bottom of the figure, our findings present the possibility of even better news on natural resources. When using the new World Bank variable to proxy for resource abundance, we find that the direct effect of resource wealth (particularly the subset of mineral resource wealth) on income growth is positive and significant. All things considered, an increase in subsoil wealth by one standard deviation—roughly the difference between Senegal and Sweden—would have brought Senegal's growth performance on a par with that of Mozambique or Kenya; not a huge improvement, but certainly not a growth curse.

Similarly, resource wealth also attenuated the risk of conflict. This is due to a positive indirect effect: Resource wealth raises income, and higher incomes, in turn, reduce the risk of conflict. Again, although the aggregate impact of resource abundance is slight—amounting to less than a 5% reduction in the risk of war in case of a standard-deviation increase in resources—it is still statistically significant. These findings are robust to using alternative measures of resource abundance, such as fuel and nonfuel mineral reserves per capita [9].

Three important caveats are relevant here. First, the number of countries in our regressions is modest (limited by the resource abundance variables). Second, consistent with most of the existing literature, our resource data do not include diamond deposits and trade flows. A focus on highly disaggregated resource measures (diamonds, but also oil) in subsequent work seems worthwhile. Third, although we believe our

resource variables represent improvements over the conventional proxy, they are not perfect. Even though differences in resource stock values are driven by differences in stocks, and not by differences in local institutions [7], they are functions of historic exploration and exploitation. Therefore, they are probably not fully exogenous. The hunt for the perfect resource variable is on, but unlikely to be settled anytime soon.

Nevertheless, our cross-country estimations cast serious doubt on the paradigm of a general resource curse. It appears as if, across the board, resource riches may be associated with higher incomes and a lower risk of civil war. Although there are undoubtedly specific countries where specific resources have eroded institutions or torn countries apart in civil strife, we find this is not the general pattern. This is consistent with several case studies that fail to show a robust link between the onset of war and resource extraction [10], and with evidence that the sector involved in turning natural resources into primary products has many more positive spillovers to the rest of the economy than often are argued [11]. Finally, it is consistent with the main message sent by the World Bank in its most recent World Development Report, which, after years of intellectual neglect, finally looks favorably at the primary sector.

The last word in the resource curse debate is far from having been spoken; but economic advisors should be aware that natural resources do not necessarily spell doom for development. Instead, their exploitation can be a valuable part of a sustainable development strategy.

References and Notes
1. J. D. Sachs, A. M. Warner, "Natural resource abundance and economic growth" (Harvard Center for International Development, Cambridge, MA, 1997).
2. P. Collier, A. Hoeffler, Oxford Econ. Paper (56), 563 (2004).
3. J. Confl. Res. 49 (special issue), (August, 2005).
4. M. L. Ross, World Politics 53, 325 (2001).
5. A. Rosser, IDS Working Paper 268 (2006); www.ids.ac.uk/ids/bookshop/wp/wp268.pdf.
6. X. Sala-i-Martin, A. Subramanian, Addressing the Natural Resource Curse: An Illustration from Nigeria (International Monetary Fund, Washington, DC, 2003).
7. World Bank, Expanding the Measure of Wealth: Indicators of Environmentally Sustainable Development (Report no. 17046, World Bank, Washington, DC, 1997).
8. W. H. Greene, Econometric Analysis (Prentice-Hall, Upper Saddle River, NJ, ed. 5, 2003), chap. 9.
9. Tables are available as supporting material on Science Online.
10. M. L. Ross, J. Peace Res. 41, 337 (2004).
11. G. Wright, J. Czelusta, Challenge 47, 6 (2004).

10.1126/science.1154539

ENGINEERING

High-Performance Transistors by Design

For large-area electronics, a redesign of the transistor structure can provide improved performance.

Xiaojun Guo and S. R. P. Silva

Consumers continue to demand more from their electronics. To fulfill needs ranging from entertainment to personalized health information in one device, the basic building block of integrated electronics, the transistor, will have to be better and cheaper.

There is growing interest in building electronics on flexible, lightweight, and cheap substrates like glass and plastics (*1*). However, in general, the deposition of semiconductor films used to make transistors on such materials has to be carried out at low temperatures to preserve substrate integrity. As a result, the quality of the organic or inorganic semiconductor films used for large-area electronics is severely constrained (*2, 3*). Disorder and impurities in the films have a dramatic influence on analog and digital performance when formed into transistors. Although approaches such as dry transfer printing (*4*) and solution casting (*5*) have been explored as ways of getting around the temperature problem, they have had little impact on mass production of high-performance devices. An alternative approach is to perform the integration at the device or circuit levels through substrate transfer techniques (*6, 7*). However, there are associated issues with this approach of higher cost and lower assembly efficiency.

In a first example of looking at transistor design, Ishii and colleagues at Hitachi (*8*) have shown that, for transistors of disordered silicon films, superior switching performance can be achieved by making the conduction channel in the transistor very thin (see the figure, left panel). A layer of nanocrystalline silicon film forms the channel and ranges from 8.0 to 2.0 nm in thickness; the precise film thickness is controlled during the deposition process. The devices have very good intrachip uniformity of electrical characteristics, even when the channel thickness is as small as 2.0 nm. For disordered silicon transistors, the performance is compromised by an anomalously high leakage current due to defect states that are present in the bandgap, which is exponentially dependent on the bandgap of the material (E_g) (*9*). When
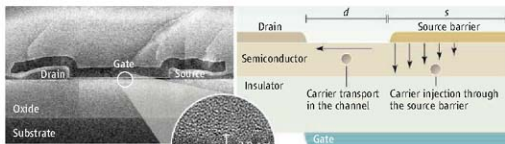
the transistor channel becomes short, the influence of the drain on the channel grows, resulting in an increased subthreshold leakage current when the gate voltage is below the threshold voltage. This subthreshold leakage is also exponentially dependent on E_g (*10*). When the silicon film thickness is reduced to a few nanometers, quantum confinement results in the modulation of the band structure, which gives rise to band-edge shifts in both the conduction and valence bands relative to bulk silicon, thereby effectively increasing E_g (*11*).

Directly attributable to this widened bandgap, the OFF-state leakage current (I_{OFF}) decreases significantly in the ultrathin channel transistors, whereas bandgap widening has a much less pronounced effect on the ON-state current (I_{ON}). The gate voltage required to switch the transistor between ON and OFF states also decreases because of a stronger gate control of the channel surface potential in thinner channel transistors (*12*). These result in a higher I_{ON}/I_{OFF} ratio, which exceeds 10^{11} for devices with a 2.0-nm-thick channel. The high I_{ON}/I_{OFF} ratio is the holy grail for designers in integrated electronics as it allows for arbitrary flexible substrates for applications where low power and fast access times are demanded and crucial.

Another example is the source-gated transistor (SGT) concept, which was first proposed for amorphous silicon in 2003 (*13*) (see figure, right panel). There is a potential barrier formed at the source contact, which governs current transport in an SGT, whereas in a regular field-effect transistor (FET), an ohmic contact is made. The gate lies under the source barrier and extends across to the drain contact. When the transistor is switched on, the current in the SGT is determined by the carrier emission over the

Structures for high-performance transistors in disordered semiconductors. (Left) Cross-sectional scanning electron microscopy and transmission electron microscopy micrographs of the fabricated nanocrystalline silicon thin-film transistors (the minimum channel thickness is 2.0 nm). (Right) Schematic of a source-gated transistor (SGT). The arrows in the channel indicate the carrier conduction in the device.

source-barrier contact, and the channel forms a parasitic resistance. A major difference between the SGT and a conventional FET is that a reverse-biased source barrier controls the current in the ON-state and current saturation is determined by the electrostatics at the source barrier rather than by pinch-off of the conduction channel at the drain. In addition, the geometry of the SGT leads to much less susceptibility to short-channel effects and a higher output impedance due to the source barrier being screened from the drain field by the gate. Compared with the FET, the SGT is operated with much less excess carrier concentration, combined with much higher internal fields over small dimensions. The low carrier concentration improves stability, whereas high internal fields increase the carrier velocity, provided the carrier velocity is proportional to the electric field (*14*). In many organic semiconductors, the mobility is field dependent. In material systems with disorder, to improve transistor speed very short channels are required to produce high internal electric fields (*15*). However, in the case for the traditional FET, achieving thin gate insulator layers of high quality to suppress the short-channel effects will be difficult, whereas the SGT can operate with very short source-drain separations even with a thick gate insulator layer, thus providing a big advantage in terms of the fabrication process. Simulation

Advanced Technology Institute, University of Surrey, Guildford, Surrey, GU2 7XH, UK. E-mail: s.silva@surrey.ac.uk

studies suggest increased carrier velocity in organic materials for high drive current and frequency response that can be orders of magnitude higher than an FET (16). The current through an SGT is insensitive to source-drain separation, thus enabling current uniformity with simple and imprecise pattern techniques.

Engineering of the transistor structure itself rather than the channel material can lead to improved device performance. It will enable the design of high-performance large-area circuits and systems based on low-cost reliable material processes, thus keeping the cost down and standard mass-production routes available to industry.

References

1. R. H. Reuss et al., Proc. IEEE. 93, 1239 (2005).
2. Y. Sun, J. A. Rogers, Adv. Mater. 19, 1897 (2007).
3. J. M. Shaw, P. F. Seidler, IBM J. Res. Dev. 45, 3 (2001).
4. Y. Sun, J. A. Rogers, Nano Lett. 4, 1953 (2004).
5. X. F. Duan et al., Nature 425, 274 (2003).
6. H.-C. Yuan, Z. Ma, Appl. Phys. Lett. 89, 212105 (2006).
7. S. A. Stauth, B. A. Parviz, Proc. Natl. Acad. Sci. U.S.A. 103, 13922 (2006).
8. T. Ishii et al., IEEE Trans. Elec. Dev. 51, 1805 (2004).
9. C. H. Kim, K.-S. Sohn, J. Jang, J. Appl. Phys. 81, 8084 (1997).
10. S. M. Sze, Physics of Semiconductor Devices (Wiley, New York, 1981).
11. Z. Lu, D. Grozea, Appl. Phys. Lett. 80, 255 (2002).
12. A. J. Walker, S. B. Herner, T. Kumar, E.-H. Chen, IEEE Trans. Elec. Dev. 51, 1856 (2004).
13. J. M. Shannon, IEEE Elec. Dev. Lett. 24, 405 (2003).
14. J. M. Shannon, F. Balon, IEEE Trans. Elec. Dev. 54, 354 (2007).
15. S. Bhattacharyya et al., Nat. Mat. 5, 19 (2006).
16. X. Guo, F. Balon, R. Hatton, J. M. Shannon, presented at 7th Annual FLEXIBLE Electronics and Displays Conference and Exhibition, January 2008, Phoenix, AZ (to be published in IEEE Special Issue on Flexible Electronics).

IMMUNOLOGY

How Frustration Leads to Inflammation

Luke A. J. O'Neill

A molecular basis for asbestos–induced inflammation may lie in a particular component of the inflammasome protein complex.

O ur bodies are constantly assaulted by infectious agents, noxious chemicals, or physical trauma. Fortunately, we have evolved a complex process—the inflammatory response—to help fight and clear the infection, remove damaging chemicals, and repair damaged tissue. The mechanisms underlying inflammation are of major interest because, as noted by British surgeon John Hunter in 1794, "when inflammation cannot accomplish that salutary purpose, it does mischief" (1). The harmful effects of inflammation can be seen in many infectious diseases, in autoinflammatory diseases such as rheumatoid arthritis, or during chronic exposure to chemicals. At worst, inflammation can provoke cancer. The mechanism by which the body senses the diverse molecular factors that cause inflammation has, until recently, been poorly understood. On page 674 in this issue, Dostert et al. (2) provide key molecular insights into how airborne pollutants, including asbestos and silica, and probably other noxious inhaled particles, lead to inflammation, pulmonary diseases, and potentially lung cancer and fibrosis.

We now know the identities of multiple cellular receptors for proinflammatory factors, notably the Toll-like receptors and NOD-like receptors, which recognize microbial products or products of damaged tissue. These receptors launch signaling cascades that induce the synthesis of proinflammatory proteins (3). Dostert et al. report that one of these NOD-like receptors, Nalp3, is required to sense the particulates asbestos and silica, and induce production of the important proinflammatory cytokine interleukin-1β. Both silica and asbestos induce the production of interleukin-1β, which likely mediates their proinflammatory effects (4, 5).

Important insights into the regulation of interleukin-1β production have come from the recent description of "inflammasomes," multiprotein complexes containing caspase-1, the enzyme that processes prointerleukin-1β into its mature form (6) (see the figure). Nalp3 (also called cryopyrin) is a component of an inflammasome that also contains the protein ASC (7). Nalp3 senses the bacterial product muramyl-dipeptide and, of greater relevance to the study by Dostert et al., monosodium urate, which causes the painful inflammatory disorder gout (8). Because both monosodium urate and asbestos are crystalline structures, the authors examined whether exposure of macrophages to physiologically relevant amounts of asbestos would induce interleukin-1β production in cells lacking Nalp3. Reducing the expression of Nalp3 attenuated caspase 1 activation and, consequently, the secretion of the cytokine.
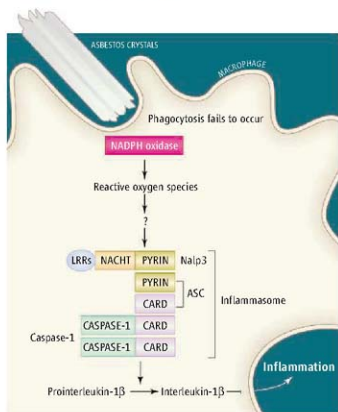
Crystal-clear inflammation. Asbestos crystals are too large to be phagocytosed by macrophages and so are subject to "frustrated" phagocytosis. This leads to activation of NADPH oxidase and the generation of reactive oxygen species. This event activates the Nalp3 inflammasome, a multiprotein complex. The protein constituents are constructed from multiple domains, as indicated. Activation of caspase-1 promotes the processing and release of the potent proinflammatory molecule interleukin-1β.

School of Biochemistry and Immunology, Trinity College, Dublin, Ireland. E-mail: laoneill@tcd.ie

As with monosodium urate, asbestos triggers an efflux of potassium from the cell. This event is required to activate Nalp3 (9), although the underlying mechanism is not known. Large crystals such as monosodium urate, or crystalline fibers such as asbestos, are subject to so-called frustrated phagocytosis and remain trapped at the cell surface, where cytoskeletal (actin) filaments form. Disruption of actin filaments with a pharmacological agent (cytochalasin D) inhibited the effect of monosodium urate and asbestos on interleukin-1β secretion. Asbestos also triggers the generation of reactive oxygen species in cells. Dostert et al. confirm this and further show that inhibitors of reactive oxygen species (such as N-acetylcysteine) block interleukin-1β production in macrophages. The source of reactive oxygen species in frustrated phagocytosis might be NADPH oxidase, an enzyme that is activated by the phagocytosis of microbes. The authors investigated a role for NADPH oxidase by using the inhibitor diphenylene iodonium, and by reducing the expression of the NADPH oxidase subunit p22^phox by RNA interference. Both approaches diminished interleukin-1β secretion in response to asbestos. Reducing the expression of thioredoxin, a protein that detoxifies reactive oxygen species, increased interleukin-1β secretion, further implicating reactive oxygen species in the

inflammatory response to this particulate.

When more mice were placed in air containing chrysotile asbestos (which is found in building materials), an increase in total cell number was observed in bronchoalveolar lavage fluid, indicative of an inflammatory reaction. By contrast, fewer cells were recruited to the lungs of Nalp3-deficient mice exposed to asbestos, and production of multiple cytokines was impaired.

Nalp3 has already been implicated in the pathological increases in interleukin-1β that occur in gout and in autoinflammatory diseases such as Muckle-Wells syndrome (10). However, precisely how Nalp3 is activated is still not clear. The current model involves the binding and hydrolysis of adenosine 5′-triphosphate to a nucleotide-binding domain of Nalp3, which is thought to lead to a conformational change in the protein, allowing activation of caspase-1 within the inflammasome (11). How reactive oxygen species affect this process is unknown. Reactive oxygen species may be particularly important for crystalline activators of the inflammasome that are subject to frustrated phagocytosis, possibly pointing to multiple mechanisms to engage with Nalp3.

A crucial finding of Dostert et al. is that Nalp3-deficient mice are resistant to asbestos-induced lung injury. An important role for interleukin-1β in this process was

already known, as was a role for this cytokine in the pathogenesis of asbestos-induced mesothelioma and in models of lung fibrosis (12, 13). The present study therefore further highlights the importance of testing the interleukin-1 receptor antagonist anakinra, which has shown efficacy in patients with other Nalp3-mediated diseases such as gout and Muckle-Wells syndrome, as a therapeutic agent to slow the progression of asbestosis and silicosis. Nalp3 itself might prove to be an interesting drug target for these diseases as well.

References

1. J. Hunter, *A Treatise on Gun-Shot Wounds, the Blood and Inflammation* (George Nicol, London, 1794).
2. C. Dostert et al., *Science* **320**, 674 (2008); published online 10 April 2008 (10.1126/science.1156995).
3. E. Creagh, L. A. J. O'Neill, *Trends Immunol.* **27**, 352 (2006).
4. Y. Ogiiso, Y. Kubota *Microbiol. Immunol.* **30**, 1189 (1986).
5. Y. Ogiiso, Y. Kubota, *Microbiol. Immunol.* **31**, 575 (1987).
6. F. Martinon, J. Tschopp, *Cell Death Differ.* **14**, 10 (2007).
7. L. Agostini et al., *Immunity* **20**, 319 (2004).
8. F. Martinon et al., *Nature* **440**, 237 (2006).
9. V. Pétrilli et al., *Cell Death Differ.* **14**, 1583 (2007).
10. M. F. McDermott, J. Tschopp, *Trends Mol. Med.* **13**, 381 (2007).
11. J. A. Duncan et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8041 (2007).
12. Y. Wang et al., *Int. J. Oncol.* **25**, 173 (2004).
13. P. F. Piguet, C. Vesin, G. E. Grau, R. C. Thompson, *Cytokine* **5**, 57 (1993).

10.1126/science.1158398

Synchronized Self-Assembly

Jeffrey S. Moore and Mary L. Kraft

Building macroscopic containers from porous membranes may be easier because of advances in controlling the kinetics of self-assembly.

Self-assembly has long been recognized as a powerful synthetic approach to obtain dynamic structures exhibiting complex functions, such as those found in nature (1). By carefully regulating non-equilibrium self-assembly, two recent studies (2, 3) demonstrate important progress, resulting in new porous membranes whose structure is controlled on several length scales.

There are two main types of self-assembly (4). Static self-assembly deals with equilibrium structures; the shapes and interaction energies of the participating components are altered to achieve various organizations. For

Kinetic control of structure. Ladet et al. successively interrupt polymer densification by removing a molded hydrogel from neutralization bath to produce multi-membrane hydrogels.

molecular components, noncovalent interactions—like hydrogen bonds, electrostatics, and van der Waals forces—are manipulated to encode building blocks with instructions that lead to the spontaneous generation of a desired target (5).

Dynamic self-assembly, on the other hand, is a non-equilibrium process in which energy is supplied to the system to maintain a steady-state population of ordered structures. Because dynamic self-assembly involves the added complexity of a sustainable

driving force, only limited progress has been made in this area (6). Recognizing that non-equilibrium self-assembly may organize matter differently from that which occurs at thermodynamic equilibrium, chemists are challenged to bring kinetic control into their repertoire of methods. This is what the two recent studies achieved.

In the first study, Ladet et al. (2) formed chitosan gels by slowly removing the water from aqueous alcohol solutions of a chitosan polyelectrolyte. This alcohol gel can be molded into shapes such as tubes and spheres of various sizes. When the gel object was bathed in a solution of an aqueous base, hydrophobic interactions within the network dominated, causing the polymer molecules to contract and form a membrane-like skin around the original object. The rate of membrane formation via densification typically

The School of Chemical Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. E-mail: jsmoore@uiuc.edu; mlkraft@uiuc.edu

occurred on the time scale of minutes. Simply removing the object from the neutralization bath interrupts membrane thickening. Insertion of the object back into the bath initiates formation of a second membrane layer between the gel-core and first membrane. Interestingly, this results in an intermembrane space that can accommodate payloads such as chondrocyte cells. Repetition of this sequence produced concentric shells of hydrogel membranes extending inward toward the core (see the first figure).

The process used to create these layered hydrogels exploits a simple type of kinetic control. Another way to drive new modes of non-equilibrium structure would be to time the release of the potential energy stored in the self-assembling components. A recent report by Capito *et al.* elegantly demonstrates such a regulated process, in which ordered structures rapidly self-assemble at the interface of two chemically distinct electrolyte solutions. This process produced functional porous membranes with complex hierarchical structures (*3*).

The authors assemble membranes instantly when two liquids come in contact, establishing a physical barrier that hinders dissipation of the ion imbalance between the small molecule electrolyte and the polyelectrolyte solutions. High osmotic pressure within the polyelectrolyte solution and the requirement for electroneutrality impels the megamolecules to extrude through pores in the structured barrier and enter the electrolyte solution. This leads to the growth of perpendicular nanofibrils in a dynamic process that is sustained by osmotic pressure (see the second figure).

The study shows that membrane-enclosed sacs of macroscopic sizes can be rapidly formed upon contact between the two liquids; moreover, the presence of polymeric electrolyte in the sac interior imparts unique self-healing function to these compartments. The resulting membranes are permeable to proteins and could be used to entrap cells by simple liquid-liquid contact to create mini-cell biology laboratories. These could then be used as controlled environments for cell expansion, stem cell differentiation, or studies of bio-sig-
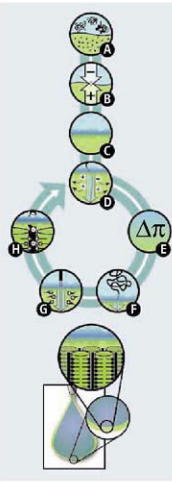
Synchronized self-assembly. In the study by Capito *et al.*, macroscopic sacs and membranes are prepared when a solution of self-assembling molecules contacts a solution of self-assembling molecules (**bottom**). Upon contact (**A**), the components are attracted by electrostatics (**B**) and instantly form a diffusion barrier (**C**). Sustained membrane growth proceeds by release of counterions (**D**), enhancing the osmotic pressure imbalance between the two liquids (**E**) that drives the polymeric electrolyte to uncoil and extrude through the barrier (**F**). New polyelectrolyte is then exposed (**G**) onto which new nanofibers assemble (**H**) causing further release of counterions (**D**).

naling from neighboring sacs entrapping colonies of other cells. By tailoring the small molecule electrolytes, the structures might be customized for a diverse array of applications in biomedicine, catalysis, and energy generation.

These recent findings point to new synthetic concepts whereby the final supramolecular structure depends on the mechanistic pathway of the assembly, rather than the thermodynamic endpoint. Ladet *et al.* show that even a relatively simple kinetic scheme can produce intriguing structures from simple components. Capito *et al.* iden-

tify the possibility of a self-sustaining pathway in which static self-assembly and a kinetically regulated mechanism combine to generate diverse architecture and functions. This finding opens the way to exciting opportunities for novel materials that may stem from incorporating pathway-directing information into the constituents and processes of self-assembly.

References
1. A. Klug, *Angew. Chem. Int. Ed.* **22**, 565 (1983).
2. S. Ladet, L. David, A. Domard, *Nature* **452**, 76 (2008).
3. R. M. Capito, H. S. Azevedo, Y. S. Velichko, A. Mata, S. I. Stupp, *Science* **319**, 1812 (2008).
4. G. M. Whitesides, B. Grzybowski, *Science* **295**, 2418 (2002).
5. G. M. Whitesides, J. P. Mathias, C. T. Seto, *Science* **254**, 1312 (1991).
6. M. Fialkowski, K. J. M. Bishop, R. Klajn, S. K. Smoukov, C. J. Campbell, B. A. Grzybowski, *J. Phys. Chem. B.* **110**, 2482 (2006).

10.1126/science.1157225

ECOLOGY

How Reefs Respond to Mass Coral Spawning

James Guest

A mass coral-spawning event at the Great Barrier Reef provided a natural experiment for studying energy and nutrient dynamics of the coral reef.

Coral spawning followed by successful larval recruitment is a crucial link in the persistence and recovery of reefs. Recent studies (*1–3*) have investigated the effects of a mass spawning event on coral reef biogeochemical processes. The research reveals how the fertilization pulse caused by spawning initiated a cascade of biogeochemical processes.

Hermaphroditic broadcast spawning is the most common reproductive strategy for reef-building corals. Broadcast spawning involves the release of buoyant, lipid-rich gametes (known as egg-sperm bundles) into the water

School of Biology, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK. E-mail: james.guest@ncl.ac.uk

column for external fertilization (see the figure). For sessile broadcast spawners such as corals, synchronous spawning within species is crucial to ensure cross-fertilization. Not only do populations exhibit synchrony, but it is common for different species to have overlapping spawning times. The first such "multispecies mass spawning event" was documented on the Great Barrier Reef (*4*), and similar events have since been witnessed in many locations (*5*). The most plausible explanation is that different coral species respond similarly but independently to timing cues and selective pressures to achieve maximum fertilization success within species (*4*, *6*, *7*).

Alternatively, spawning at the same time may saturate predators. Embryos and larvae

drifting across the reef are met by a "wall of mouths" (8–10); spawning together may increase the chance that at least some progeny will avoid being eaten. Consumption converts gametes to fecal matter, which sinks to the sea floor. Spawn-derived organic matter may also reach the sea floor via microbial breakdown and direct sedimentation of unfertilized eggs or by introduction of diluted sperm. Spawn slicks (see the figure) may be transported away from natal reefs, with implications for reef connectivity and spatial variation in reef productivity. In lagoons, slicks can become trapped on the sediment surface, resulting in a large input of carbon, nitrogen,

example, particulate nitrogen concentrations in the water column peaked after spawning and stayed high for 17 days, and the concentration of particulate organic matter reaching the sediment was elevated for 2 weeks. Isotopic signatures of the suspended matter reflected that of coral gametes, revealing that organic matter from coral spawn was immediately transferred to the food web. These findings show how the reef ecosystem responds to a sudden increase in nutrient load and rapidly processes the large organic matter pulse from coral spawning.

The spawning event stimulated biological activity both at the sea floor and in the water

limited due to the large reservoir of bioavailable phosphorus stored in the top few centimeters of the reef sediment.

Further research is needed to examine and compare the response of reefs to mass spawning among and within regions. For example, many ecological differences exist between Indo-Pacific and Atlantic reefs. Compared to most Indo-Pacific reefs, in the western Atlantic there are a higher proportion of corals that internally "brood" larvae versus corals that broadcast spawn; nonetheless, multi-species spawning events involving fewer species are well documented on some western Atlantic reefs (13).
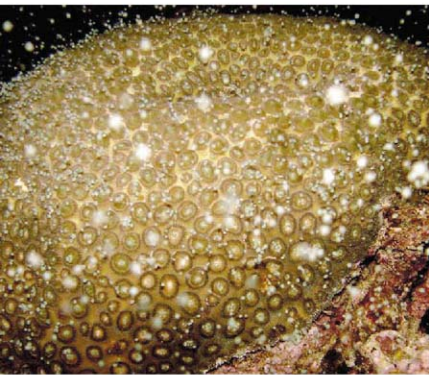
Coral spawning and spawn slick. (Left) A coral releasing buoyant egg-sperm bundles at night. (Right) Spawn slicks—aggregations of buoyant eggs, embryos, and larvae—like this one at Scott Reef, Western Australia, are often seen the day after spawning.

and phosphorus. On rare occasions, this has led to mass mortalities of reef fauna due to extremely large reductions in oxygen concentrations (11).

A small spawning event at Heron Island (Great Barrier Reef) in 2001 was shown to stimulate sedimentary oxygen consumption rates for up to 9 days after spawning (12). Clearly, degradable organic matter from the spawning was enhancing metabolism on the reef. These links between coral spawning and the energy and nutrient dynamics of the coral reef system prompted Eyre et al., Glud et al., and Wild et al. to examine a major mass spawning event at Heron Island in November 2005. The event had an immediate but short-term effect on the concentration of particulate organic matter detected in the water column and the sediments. For

column. For example, biological oxygen demand—a measure of how fast organisms use up oxygen—and chlorophyll a concentrations in the water column increased substantially after the spawning. Mass-balance calculations indicated an efficient mineralization of spawn-derived nitrogen by microbial communities within the reef sands. Biological activity was consistently higher near the sea floor than in the open ocean, highlighting the central role of reef sedimentary metabolism in processing organic matter on the reef.

The postspawning phytoplankton bloom removed almost all the dissolved inorganic nitrogen, whereas only a small proportion of the dissolved inorganic phosphorus was removed. This suggests that pelagic primary production is nitrogen limited. Benthic primary production is also likely to be nitrogen

The latest research on mass spawning (1–3) highlights the importance of examining the response to a large nutrient addition at an ecosystem scale. The results reveal a strong link between the organic materials contained in coral spawn and reef primary production. Central to the nutrient recycling process were the highly permeable reef sands and associated assemblages of heterotrophic prokaryotes. This biocatalytical filter system (14) allows reefs with naturally low background nutrient concentrations to buffer extreme nutrient peaks. The reef sands thus have an important function for reef resilience. Benthic algal overgrowth—reported from many reef locations worldwide—may compromise sediment function and reduce resilience.

Coral reefs are increasingly exposed to high nutrient levels as a result of human activ-

ities. The nutrient pulse from spawning is efficiently trapped, but this efficiency may become a threat on reefs exposed to continuous coastal eutrophication. Reefs within the Great Barrier Reef occur along gradients of water quality depending on their distances from point sources of pollution. Comparisons of ecosystem responses to spawning among these reefs may help to elucidate how these systems function under different levels of stress, as well as providing greater insights

into coral reef functioning and resilience.

References

1. B. D. Eyre, R. N. Glud, N. Patten, *Limnol. Oceanogr.* **53**, 997 (2008).
2. R. N. Glud, B. D. Eyre, N. Patten, *Limnol. Oceanogr.* **53**, 1014 (2008).
3. C. Wild, C. Jantzen, U. Struck, O. Hoegh-Guldberg, M. Huettel, *Coral Reefs* **27**, 123 (2008).
4. P. L. Harrison *et al.*, *Science* **223**, 1186 (1984).
5. J. R. Guest, A. H. Baird, B. P. L. Goh, L. M. Chou, *Invertebr. Reprod. Dev.* **48**, 207 (2005).
6. R. C. Babcock *et al.*, *Mar. Biol.* **90**, 379 (1986).
7. J. K. Oliver, R. C. Babcock, P. L. Harrison, B. L. Willis, *Proc. 6th Int. Coral Reef Symp.* **2**, 803 (1988).
8. W. M. Hammer, M. S. Jones, J. H. Carleton, I. R. Hauri, D. M. Williams, *Bull. Mar. Sci.* **42**, 459 (1988).
9. A. H. Baird, M. S. Pratchett, D. J. Gibson, N. Koziumi, C. P. Marquis, *Mar. Freshw. Res.* **52**, 865 (2001).
10. M. S. Pratchett, N. Gust, G. Goby, S. O. Klanten, *Coral Reefs* **20**, 13 (2001).
11. C. J. Simpson, J. L. Cary, R. J. Masini, *Coral Reefs* **12**, 185 (1993).
12. C. Wild, R. Tollrian, M. Huettel, *Mar. Ecol. Prog. Ser.* **271**, 159 (2004).
13. C. Wild *et al.*, *Nature* **428**, 66 (2004).
14. D. K. Hagman, R. S. Gittings, J. P. Deslarzes, *Gulf Mex. Sci.* **2**, 170 (1998).

PHYSICS

A Unified Picture of Laser Physics

Diffusive random lasers, whose mechanism has been elusive, are now explained by a general theory that encompasses conventional lasers.

Jorge Bravo-Abad and Marin Soljačić

Laser technology is present in our daily lives through literally thousands of applications, including surgical instruments, CD and DVD players, optical fiber communications, and even supermarket bar-code readers. Despite the fast pace of laser research, the design of most laser devices relies on assumptions in the underlying theory that have barely changed since the early days of laser theory (*1*). However, this situation is problematic for two reasons. First, the rapid advance of nanofabrication techniques has led to the development of completely new lasing systems whose description falls outside the scope of conventional laser theory. Of these, random lasers (*2*) are perhaps the most challenging example. Second, more general models could enable the design of substantially different classes of lasers. With their contribution on page 643 of this issue, Türeci *et al.* (*3*) have substantially changed this picture. By developing a new theory in which the main properties of a laser can be physically understood as the result of strong nonlinear interactions between lasing modes, they have provided a substantially broader perspective of laser physics that unifies the physical description of many possible laser structures.

The most common description of a laser is that of an active lasing material, or gain medium (which could be an atomic vapor, a solid, or a dye), inside a resonant cavity formed by two mirrors. If the lasing material is properly pumped by an external excitation (which can be optical, electrical, etc.), most of the basic constituents of the lasing media (such as atoms, molecules, or ions) will be in

excited states—it is said that the population inversion condition has been reached (*1*). Then, light of a certain frequency propagating through this medium will stimulate emission of radiation of the same frequency (and same direction) from the excited states. This process creates an amplifying medium: Light will be coherently amplified as it bounces back and forth inside the cavity, producing an output beam that is both highly directional and monochromatic (see the figure, left panel).

Standard laser theory explains the physics behind these devices, which resemble Fabry-Perot etalons and interferometers, provided that light is tightly trapped inside the cavity—that is, when most of the energy of the amplified light remains inside the cavity for a long time (*1*). It also assumes that both the corresponding lasing frequencies and lasing modes are essentially determined by the modes and frequencies of the resonant cavity when the lasing material is not present.

The universality of this description has recently been challenged with the appearance of the so-called diffusive random lasers (*4*). A random laser consists essentially of a set of

particles that scatter light and are embedded in a gain medium (see the figure, right panel). Most random lasers operate in the so-called diffusive regime, in which there are no light-confinement mechanisms in the absence of the amplifying medium. Thus, diffusive random lasers apparently lack the strong light-trapping mechanism that conventional laser theory regards as an essential ingredient for efficient light amplification.

At first glance, one might think that the physical mechanisms responsible for the lasing action in conventional and diffusive random lasers are very different. However, it has been shown experimentally that these two systems have the same basic features (*4*–*7*), which seems to suggest that a unified description should be possible. Until now, the theoretical analysis of random lasers has been restricted to fully numerical approaches (*8*), which, although interesting, do not offer a fundamental physical insight into mechanisms.

The theory developed by Türeci *et al.* to describe lasers provides the missing physical insight in an intrinsically elegant manner. By substituting the role of linear cavity reso-
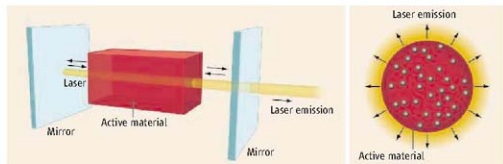
Two different lasers, just one theory. (Left) Sketch of a conventional laser: Light is tightly confined between two mirrors that define the resonant cavity modes and laser frequency. **(Right)** A random laser: Light is scattered by a set of particles embedded in an active material. Türeci *et al.*'s theory provides a unified description of the physics behind both this system and the conventional laser shown in the left panel.

Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: jbravo@mit.edu

nances with a new set of modes—the constant-flux states—the authors find a simple analytical expression from which all of the properties of any laser structure can be obtained, given a knowledge of the dielectric constant profile of the system together with the main parameters characterizing the amplifying material (such as the amplification profile or the atomic frequency).

The versatility of their approach is demonstrated by applying it to the debate about the physical basis of lasing in diffusive random lasers. They show that for these kinds of structures, the lasing frequency predicted by standard laser theory is substantially modified by a new contribution that has no analog in conventional lasers. In addition, the theoretical framework developed by Türeci et al. allows us to track the competition between lasing modes within these systems as a result of a strong nonlinear interaction through the gain medium.

More specifically, the authors show how pairs of modes of nearly identical frequencies compete with each other in a complex manner that ultimately determines both the emitted frequencies of the diffusive random lasers and their corresponding intensities; these predictions are in very good agreement with the recent experimental results (7). Finally, this new perspective on lasers reveals how the electric field profile in a diffusive random laser is more intense at the edge of the system than anywhere else in the sample—an interesting property that has also been observed in conventional lasers in which light is weakly confined.

In addition to its importance for understanding most of the physical properties discovered recently in diffusive random lasers, the theory of lasing provided by Türeci et al. may inspire the design of substantially different classes of structures that could be the basis

of improved laser-based devices. From a more fundamental standpoint, this work could spark a new branch of nonlinear dynamics, in which phenomena such as optical bistability or multistability could be explored in novel types of lasing structures.

References

1. A. Siegman, *Lasers* (University Science, Mill Valley, CA, 1986).
2. N. M. Lawandy, R. M. Balachandran, A. S. L. Gomes, E. Sauvain, *Nature* **368**, 436 (1994).
3. H. E. Türeci, L. Ge, S. Rotter, A. D. Stone, *Science* **320**, 643 (2008).
4. H. Cao, *J. Phys. A* **38**, 10497 (2005).
5. H. Cao et al., *Phys. Rev. Lett.* **82**, 2278 (1999).
6. S. Mujumdar, M. Ricci, R. Torre, D. S. Wiersma, *Phys. Rev. Lett.* **93**, 053903 (2004).
7. K. L. van der Molen, R. W. Tjerkstra, A. P. Mosk, A. Lagendijk, *Phys. Rev. Lett.* **98**, 143901 (2007).
8. C. Vanneste, P. Sebbah, H. Cao, *Phys. Rev. Lett.* **98**, 143902 (2007).

MOLECULAR BIOLOGY

The Paradox of Silent Heterochromatin

Ingela Djupedal and Karl Ekwall

Cell cycle control of heterochromatin disassembly may explain the paradox of heterochromatin gene expression.

E ukaryotic chromosomes are generally partitioned into euchromatin and heterochromatin. The former is associated with actively expressed genes, whereas the latter has been considered to be inaccessible to RNA polymerase II, the enzyme that transcribes DNA to RNA. Heterochromatin is therefore transcriptionally silent. Paradoxically, to remain silent, heterochromatin that is located at a chromosome's centromere—a region that is essential for chromosome separation during cell division—depends on both its transcription by RNA polymerase II and inhibition of its transcription by a mechanism called RNA interference (1–3). Two recent studies, by Kloc et al. (4) and by Chen et al. (5), demonstrate that transcription of this heterochromatin depends on the stage of the cell division cycle and thereby provides a possible solution to the paradox. According to the proposed model, the heterochromatic structure breaks up at the M phase (mitosis) of the cycle, enabling its transcription by RNA polymerase II, with subsequent reassembly

of heterochromatin and gene silencing.

The cell cycle of a eukaryotic cell is divided into four distinct phases: DNA replication at S phase; growth and preparation for division in G_2; nuclear division at M phase; and G_1, during which the cell may exit the cell cycle or continue dividing. Previously, heterochromatin has only been studied in asynchronous growing cell cultures, and cell cycle–dependent effects have not been appreciated. Kloc et al. and Chen et al. used synchronized cells to investigate the cell cycle dynamics of transcription and heterochromatin assembly at centromeres in the fission yeast *Schizosaccharomyces pombe*, the model organism in which the involvement of RNA interference in heterochromatin formation was first described (2). Both groups found that transcripts corresponding to centromeres (where there are large numbers of repeated DNA sequences) accumulate at S phase. This coincides with increased amounts of small interfering RNAs (siRNAs), which are derived from these transcripts. These siRNAs in turn silence the transcription of RNA from the centromeres by RNA interference. The assembly of heterochromatin at centromeres is also cell cycle dependent. Typical heterochromatic marks, such as the methylation of a lysine residue on

histone 3 (H3K9me); histones are the major protein constituents of chromatin) and binding of the heterochromatin protein Swi6, decreased at S phase, causing the heterochromatin structure to become more loosely packed. Simultaneously, other marks that are associated with actively transcribed genes were detected. At the onset of M phase and remaining throughout the following S phase, there was a peak of phosphorylation of serine 10 on histone H3 (H3S10ph)—a mark antagonistic to Swi6 binding (6, 7)—together with methylation of lysine 4 on histone H3 (H3K4me). These changes further indicate that heterochromatin structure is changing and becoming more permissive to transcription.

These results support a stepwise model of cell cycle–regulated reassembly of heterochromatin at each cell division (see the figure). The densely packed structure of heterochromatin is dissolved at mitosis, followed by the binding of RNA polymerase II at S phase (5) and transcription. The transcripts are processed into siRNA that together with the RNA interference machinery directs the formation of heterochromatin to loci that are complementary to the siRNA. In addition, RNA polymerase II both directly and indirectly interacts (the latter, via nascent RNA)

Department of Biosciences and Medical Nutrition, Karolinska Institutet, Sweden/School of Life Sciences, University College Södertörn, NOVUM, 141 57 Huddinge, Sweden. E-mail: karl.ekwall@ki.se

- RNA interference machinery
- Lysine methyltransferase (KMT)
- Histone deacetylase (HDAC)
- Nucleosome
- Swi6
- H3K9me
- H3S10ph
- H3K4me

Loosely packed heterochromatin
RNA synthesis
siRNA production and RNA interference

Densely packed heterochromatin
No transcription?
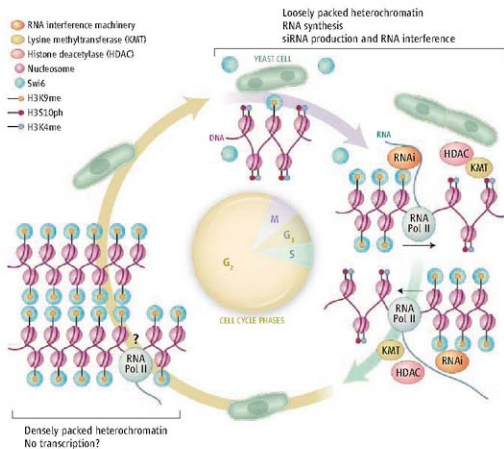
Cell cycle regulation of heterochromatin assembly. A G$_2$ cell (*S. pombe*) has a single nucleus and elongates until it reaches M phase (mitosis). Heterochromatin located at chromosomal centromeres becomes differentially methylated and phosphorylated on histones throughout the cell cycle as indicated. These modifications control the binding of the heterochromatin protein Swi6. During DNA replication (S phase), a more accessible chromatin structure permits RNA polymerase II to transcribe centromeric DNA, which in turn recruits the RNA interference machinery to silence the expression of the same loci.

with proteins that form heterochromatin. These proteins include histone deacetylases and lysine methyltransferases responsible for the heterochromatin silencing marks that are detected after the onset of S-phase transcription (*5*). Further supporting the model, phosphorylation of H3S10 was shown by Kloc *et al.* to be required for maintaining methylation of H3K9 in successive cell divisions (*4*).

There may be one important question that remains to be answered to solve the paradox that transcription is necessary to establish transcriptionally silent heterochromatin. Both Kloc *et al.* and Chen *et al.* show a relative increase of heterochromatin transcription in S phase, but is there any low-level constitutive transcription throughout the rest of the cell cycle? In certain mutants of *S. pombe*, both strands of double-stranded DNA at centromeres are transcribed (*2*). In wild-type cells, the "forward" transcript (RNA that is synthesized from one of the DNA strands) is silenced by heterochromatin, whereas transcripts from the "reverse" strand (RNA synthesized from the complementary DNA strand but in the opposite direction) are easily detected in

unsynchronized cells, of which most (>70%) are in the G$_2$ phase. Kloc *et al.* propose that transcription of both strands preferentially occurs in S phase, after DNA replication, when the heterochromatic structure is more permissive, presenting a solution to the paradox of transcription of silent heterochromatin. Only a relative representation of transcripts in G$_2$ is shown, however, making it difficult to assess absolute transcript levels (*4*). Chen *et al.* report constitutive transcription of the reverse strand at each time point investigated (*5*). However, these results must be considered uncertain because the cells were synchronized by a shift to restrictive temperatures, and Kloc *et al.* nicely demonstrate that RNA interference is inhibited at such elevated temperatures (*4*). The reported temperature dependency of RNA interference is in itself interesting and may explain observations from several organisms where gene silencing is suppressed at high temperatures but enhanced at low temperatures (*8–10*). Vernalization in plants—the induction of growth and flowering by exposure to low temperature—is one example in which the increased activity of RNA interference dur-

ing winter would enhance gene silencing, thus providing means for organisms to respond to environmental cues.

The work by Kloc *et al.* and Chen *et al.* presents an elegant mechanism for the inheritance of epigenetic mechanisms of gene regulation from mother to daughter cells—that is, the inheritance of histone modifications that control gene expression through the reassembly of heterochromatin at the beginning of each cell cycle. A similar mechanism has recently been described, in which cell cycle–dependent readthrough transcription (when RNA synthesis continues past a termination signal encoded in the DNA) is coupled to formation of transient heterochromatin at active, convergent genes (which are transcribed in opposite directions toward each other) (*11*). Cell cycle dynamics of heterochromatin and transcription may therefore represent a general regulatory mechanism. If the reverse DNA strand of centromeres is constitutively transcribed at low levels in *S. pombe*, the regulation may be akin to that of several known examples of genes residing within heterochromatin (*12*), many of which depend on this structure for their expression. In the fly *Drosophila melanogaster*, the actively transcribed genes *light* and *chinitase* have the heterochromatic mark H3K9me (*13*). Whether this type of transcription is mechanistically different from the transcription of euchromatic genes remains to be determined. The study by Kloc *et al.*, in combination with other recent studies, ascribes a more dynamic function of heterochromatin than previously anticipated (*4, 5, 11, 13*), and it is clear that the simplified view of heterochromatin as a structure inaccessible to RNA polymerase II should be revised.

References

1. I. Djupedal *et al.*, *Genes Dev.* **19**, 2301 (2005).
2. T. A. Volpe *et al.*, *Science* **297**, 1833 (2002).
3. H. Kato *et al.*, *Science* **309**, 467 (2005).
4. M. Z. Kloc, M. Zaratiegui, E. Nora, R. Martienssen, *Curr. Biol.* **18**, 490 (2008).
5. E. S. Chen *et al.*, *Nature* **451**, 734 (2008).
6. W. Fischle *et al.*, *Nature* **438**, 1116 (2005).
7. T. Hirota, J. J. Lipp, B. H. Toh, J. M. Peters, *Nature* **438**, 1176 (2005).
8. Z. J. Chen, L. Tian, *Biochim. Biophys. Acta* **1769**, 295 (2007).
9. R. C. Allshire, J. P. Javerzat, N. J. Redhead, G. Cranston, *Cell* **76**, 157 (1994).
10. V. K. Lloyd, D. Dyment, D. A. Sinclair, T. A. Grigliatti, *Genome* **46**, 1104 (2003).
11. M. Gullerova, N. J. Proudfoot, *Cell* **132**, 983 (2008).
12. J. C. Yasuhara, B. T. Wakimoto, *Trends Genet.* **22**, 330 (2006).
13. J. C. Yasuhara, B. T. Wakimoto, *PLoS Genet.* **4**, e16 (2008).

Structure and Dynamics of Earth's Lower Mantle

Edward J. Garnero* and Allen K. McNamara

Processes within the lowest several hundred kilometers of Earth's rocky mantle play a critical role in the evolution of the planet. Understanding Earth's lower mantle requires putting recent seismic and mineral physics discoveries into a self-consistent, geodynamically feasible context. Two nearly antipodal large low-shear-velocity provinces in the deep mantle likely represent chemically distinct and denser material. High-resolution seismological studies have revealed laterally varying seismic velocity discontinuities in the deepest few hundred kilometers, consistent with a phase transition from perovskite to post-perovskite. In the deepest tens of kilometers of the mantle, isolated pockets of ultralow seismic velocities may denote Earth's deepest magma chamber.

Earth's most profound internal boundary lies roughly halfway to its center, at a depth of nearly 2900 km, where the solid mantle meets the fluid outer core. Emerging research characterizes structure and processes on the mantle side of this boundary that influence chemistry and convection throughout the mantle, heat loss from the core, and Earth's thermal structure and evolution. The fields of seismology, mineral physics, geodynamics, and geochemistry have been key in providing new information. To better understand Earth's lowermost mantle, and hence whole-mantle processes, we summarize several recent observations and examine them in a geodynamical context.

Historically, the lowermost few hundred kilometers of the mantle was noted as having a reduced seismic velocity gradient with depth, interpreted as being caused by a lowermost-mantle thermal boundary layer above the hot core. This view was modified in the early 1980s, when seismologists observed a first-order discontinuous increase in velocity between 250 and 350 km above the core-mantle boundary (CMB) (1). This discontinuous jump is typically referred to as the D'' discontinuity. In the past few years, observations, modeling, and predictions have shown that the deepest mantle is complex (Fig. 1) and much more anomalous than the rest of the lower mantle. The term D'' is used to refer to the general depth shell of the lowermost several hundred kilometers of the mantle, and does not denote any specific structural characteristic.

Long-Wavelength Heterogeneity and Implications

Earth's internal structure is predominantly imaged by seismic methods. Tomographic inversions of seismic data yield maps of seismic-wave speed heterogeneity throughout the mantle, with a resolution typically greater than 1000 km (Fig. 1 and Fig. 2A). It has long been known that convergent

School of Earth and Space Exploration, Arizona State University, Box 871404, Tempe, AZ 85287, USA.

*To whom correspondence should be addressed. E-mail: garnero@asu.edu

plate boundaries overlie regions of D'' with higher than average velocities, and that hot-spot volcanoes overlie regions with lower than average velocities. Such spatial correlations, with evidence for high P- and S-wave velocities mimicking slab shapes extending from beneath some subduction zones well into the lower mantle, constitute one argument in favor of whole-mantle convection (2, 3).

Seismic data suggest that two broad regions with lowered shear-wave speeds and higher than average density lie beneath the Pacific and Africa (4, 5). The African anomaly appears to extend upward from the CMB about 1000 km (6), whereas the height of the Pacific anomaly is less certain but probably at least 400 to 500 km. Each anomaly is about 15,000 km across, and together they cover nearly 50% of the CMB. The boundaries between these large low-shear-velocity provinces (LLSVPs) and normal mantle are sharp, as implied by seismic waves that graze LLSVP edges (6–8).

Geodynamic calculations show that chemically distinct deep-mantle regions can be formed and maintained, whereby mantle convection sweeps intrinsically denser material toward underlying regions (9, 10) resembling LLSVP geometry (Fig. 2, A and B). This material must have intrinsic density elevated by a few percent relative to the surrounding mantle; if higher, structures would flatten out along the CMB, and if lower, they would be easily entrained in upwellings. If the resultant density is less than that of the surrounding mantle, the material is buoyant and forms large, doming structures that actively rise through the mantle (Fig. 2C). Alternatively, these thermochemical superplumes may heat up and rise because of excess thermal buoyancy, then cool and sink because of decreased thermal buoyancy (11). Smaller plumes that entrain some of the denser material can form on the tops of these structures. Assuming this behavior for Earth implies that LLSVPs are superplumes in various stages of ascent or fall (11).

If LLSVP thermal and compositional buoyancy are roughly balanced, then near-neutral or negative buoyancy can yield long-lived stable structures (9, 10, 12, 13). Piles are passively swept and shaped by mantle convection, preferentially accumulating into ridge-like structures beneath dominant upwelling centers (the Pacific and Africa). Plumes may originate from pile tops, particularly at peaks or ridges (Fig. 2B), and entrain nominal amounts of pile material. If the bulk modulus of pile material is anomalously high (e.g., consistent with subducted basaltic crust), structures intermediate to piles and superplumes can form, which have steeper sides than the other cases (13).

Lower-mantle chemical heterogeneity might also have existed since Earth's early differentiation. Seismic images of present-day structure would
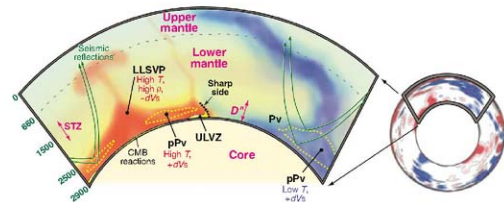
Fig. 1. Tomographically derived (43) high and low seismic shear velocity variations in Earth's mantle (blue and red, respectively) are shown in an equatorial cross section (right) viewed from the south, along with an enlarged panel (left) depicting several seismic findings in the D'' region. A large low-shear-velocity province (LLSVP) is found beneath the Pacific Ocean and Africa, and has high density (ρ) and temperature (T), with sharp sides between LLSVP low velocities ($-dVs$) and surrounding mantle. A pair of seismic reflectors seen beneath subduction as well as within the LLSVP is consistent with a double crossing of the perovskite (Pv) to post-perovskite (pPv) phase transition (yellow dashed lines), which locally elevates shear velocity ($+dVs$). Ultralow-velocity zone (ULVZ, yellow) material sits atop the CMB and can be swept around in lateral currents, possibly relating to chemical reactions between the mantle and core. The spin transition zone (STZ) centered near 1500 km depth represents a change in the spin state of Fe^{2+} to Fe^{3+} and may also affect lower-mantle densities and velocities (44, 45).

thus depict remnants of this material, which may be slowly entraining away through time. Alternatively, denser material might have accumulated throughout Earth's history and might consist of chemical reaction products from the CMB (14) or subducted oceanic crust (13, 15, 16). Piles composed of a long-lived primordial layer will likely have sharp contacts at their top surface (Fig. 2B), whereas those composed of accumulated subducted crust may have a rough or diffusive top (Fig. 2D).

Geodynamic models predict that superplumes have doming, concave-downward tops; piles, on the other hand, have concave-upward tops surrounded by ridges. A velocity reduction some 300 to 400 km above the CMB is imaged in the central Pacific (17) and may represent the top of the Pacific pile. Improving the resolution of the

Fig. 2. (A) Shear velocity perturbations between 660 km depth and the CMB, isocontoured at ±0.6% (blue/red) for model S20RTS (43). Sharp LLSVP edges (yellow dashed lines) have been noted for both LLSVPs (8). **(B to D)** Geodynamical modeling results illustrating different conceptual models to explain LLSVPs observed beneath Africa and the Pacific (see supporting online text for details of calculations). All calculations extend from the CMB to the surface, and the 3D calculations are performed in spherical geometry, unwrapped into a box for easier visualization. **(B)** 3D thermochemical piles (gold), shown with a whole-mantle temperature cross section (red and blue denote hot and cold, respectively). **(C)** 3D thermochemical superplumes (gold). **(D)** 2D calculations of transient thermochemical piles created by accumulated subducted crust, with temperature and composition superimposed (red/blue = hot/cold; denser material is darker).

sharpness and variability of this reduction will assist in constraining LLSVP origin and evolution.

The trace chemistry of ocean-island basalts (OIBs), if from a deep-mantle origin, may shed light on LLSVP chemistry. Volcanic hot spots tend to overlie LLSVP edges rather than their interiors (18), which is consistent with edges and ridges of thermochemical piles forming in regions of return flow and initiating plumes (9, 10). Efforts to directly image plumes are encouraging (19) but controversial (20). Numerical models of mantle convection show that plume morphologies are often more complicated than simple vertically continuous whole-mantle conduits (e.g., bent and curved central conduits, complicated plume head shapes, thin or diminishing plume tails, etc.).

Regional D″ Studies

Evidence for a shear-wave velocity discontinuity at the top of D″ is unequivocal. It is

South America, and dense seismic networks in North America.

Typically, attributes of seismic-wave reflectance are extracted throughout some volume of interest by stacking or summing seismic records. These data permit mapping of lateral variations in D″ discontinuity height above the CMB. For example, a 100-km vertical offset in D″ discontinuity depth is observed beneath the Cocos plate (21, 22). A second, deeper reflector is evident in some studies beneath the velocity increase, mapped as a velocity reduction (21–23). Beneath the central Pacific, up to four discontinuities in the shear velocity have been detected (17). A major uncertainty in interpreting the seismic data is the assumption of a reference velocity structure. For example, lateral variations in the height of the D″ discontinuity, underlain by constant velocity, can trade off with a model containing a fixed height but underlain with variations in velocity. Studies
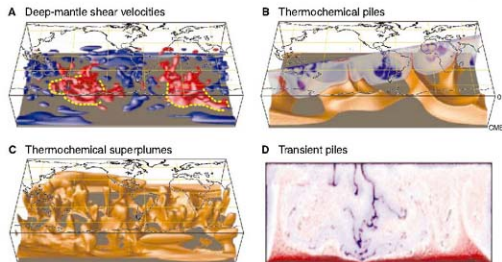
of P waves have been less common than for S and typically lack clear waveform evidence for D″ reflections, although there are some exceptions (24). More commonly, stacking of hundreds or thousands of seismograms is necessary to detect reflected P-wave energy (25); thus, the discontinuity is a much weaker reflector for P waves.

Lateral variations in deep-mantle temperature are expected but should be smooth, and hence they do not explain a step velocity increase associated with the D″ discontinuity, nor differing S- and P-wave discontinuity strengths. Furthermore, numerical models predict that lowermost-mantle thermal boundary layers should be thin, much less than the observed discontinuity height. Thus, D″ has been variously interpreted as chemical dregs from subduction, as a region of chemical reaction between the core and mantle, as a boundary between isotropic and anisotropic fabrics, or as a solid-state phase change (26). Recently, the latter hypothesis has gained favor, owing to the recent discovery of an exothermic phase transition in deep-mantle perovskite (Pv) to a post-perovskite (pPv) structure (27–29).

If this phase transition is responsible for the D″ discontinuity in cold, high-velocity regions, then the experimentally determined positive Clapeyron slope of the transition predicts that the discontinuity should deepen or even vanish in hot areas. However, clear evidence is present for an S-wave discontinuity within the Pacific LLSVP (17), with height above the CMB and topographical variability both similar to the region beneath the Cocos Plate. This supports the idea that LLSVPs may have a different mantle composition (perhaps increased iron abundance) that perturbs the phase boundary depth. However, some form of chemical stratification could also generate reflections in the LLSVP.

Because temperature is expected to strongly increase with depth near the hotter outer core, post-perovskite may revert back to perovskite closer to the CMB, depending on Clapeyron slope and temperature. Thus, a lens of post-perovskite is possible from a double crossing of the phase boundary (30). Attributing a pair of seismically detected D″ discontinuities (i.e., a velocity increase overlying a decrease) to post-perovskite permits determination of temperature at the two discontinuity depths, if thermal conductivity and Clapeyron slope are known. In turn, this permits estimation of core heat flux of 50 to 100 mW/m² (17, 23), a range of values much higher than estimates based on plume buoyancy flux. Although the phase transition is not expected to greatly alter first-order dynamics, an exothermic phase change at these depths could enhance the generation of plumes, affecting CMB heat flow (31). A lower CMB temperature would preclude the double crossing.

Accurate characterization of the deeper discontinuity is challenging because P- and S-wave reflections from a discontinuous velocity reduction at the magnitude predicted by mineral physics

studies (32) should be weak (33). Also, small-scale heterogeneity may complicate the detection of the lower boundary by scattering energy (22). Another challenge is that the P-wave jump associated with the entrance into the ppV phase is predicted to be small and negligible—fractions of a percent, and possibly even negative (32).

Horizontal and vertical components of shear waves with appreciable path lengths in the deepest mantle have slightly different arrival times for some regions, indicating seismic anisotropy in D″. Anisotropy is detected beneath implied down-welling flow (e.g., beneath the Caribbean, Alaska, and Eurasia) and upwelling (e.g., beneath the central Pacific) and has been suggested to be weak outside these regions (34, 35). Most data imply that the orientation of anisotropy is three-dimensional, with shear-wave splitting depending on azimuth. If the link between observations and dynamical predictions can be properly established, it may be possible to provide constraints on mantle flow patterns and rheology of the lower-most mantle (36). The main minerals in the lower mantle—Pv, ppV, and magnesiowüstite—are each strongly anisotropic but may respond differently to deformation, which makes inference of mantle flow patterns from observed anisotropy more difficult.

As Pv transforms to ppV, preexisting fabric undergoing transition is expected to be severely altered, and perhaps erased, by the major change in crystal structure; however, preexisting fabrics in magnesiowüstite should remain and combine with developing fabric in the new ppV grains at greater depths. An additional complication is that magnesiowüstite and ppV likely have different rheologies, such that the weaker of the two minerals may accommodate much more deformation [hence, more lattice-preferred orientation (LPO)] than the stronger mineral. If ppV is the cause of the D″ discontinuity and is also the dominant mineral associated with D″ anisotropy, then there may be an offset between the depths of the discontinuity and the onset of anisotropy, because some finite amount of deformation is required to develop LPO in ppV (fig. S3). This may explain seismic observations under the central Atlantic, thought to be away from current downwellings, in which there is evidence for a D″ discontinuity (37) but only weak (or absent) seismic anisotropy (35).

Ultralow-Velocity Zones

Thin patches (5 to 40 km thick) in which P- and S-wave velocities are reduced by up to 10% and 30%, respectively, have been imaged directly above the CMB in several places (38). These characteristics are consistent with the presence of some melt in the lowermost mantle (39). An ultralow-velocity zone (ULVZ) density increase of up to 10% has also been noted (40). If the origin of the ULVZs is partial melt, they should be present wherever rock of appropriate composition exists, because the CMB is isothermal. However, ULVZs should be thickest in the hottest regions above the CMB. Modeling implies that these regions are at the base of

plumes in an isochemical mantle, which would correspond to the center of LLSVPs (fig. S4A). However, if LLSVPs are thermochemical structures, the hottest regions should be at their edges (8, 9, 13). Several studies show evidence for ULVZ structure near the edge of the Pacific LLSVP (17, 40), consistent with it being a thermochemical structure (fig. S4B). Assessing the geometry and extent of ULVZs is challenging because their thickness may approach the limits of seismic detection. Thin ULVZs having only mild velocity drops are even harder to detect.

Other mechanisms that produce ULVZ material are chemical heterogeneity from the mantle chemically reacting with the core (14) and CMB underside sedimentation (41). However, the 3-to-1 ratio of S-to-P velocity reduction likely requires some degree of partial melting. An interesting possibility involves ppV: the deepest mantle is cool enough to avoid a double crossing of the ppV phase boundary, even in iron-rich ppV can explain an ULVZ layer (42). However, recent work imaged both a double crossing of the ppV phase boundary and a deeper ULVZ layer in the Pacific (14). For a chemical and/or partial-melt origin, episodic entrainment of ULVZ material may be possible (8), leading to heterogeneities throughout the D″ layer in certain regions. Viscous convective forces may be strong enough that even high-density (>10%) material can be swept either to the base of plumes or to the edges of chemically distinct piles.

Next Steps

Seismological studies reveal a lowermost mantle rich in complexity on a variety of spatial scales: the large LLVSPs beneath Africa and the Pacific, the regionally variable D″ discontinuities observed ~200 to 300 km above the CMB, and isolated ULVZs in the lowermost tens of kilometers of the mantle. Mineral physics experiments imply that heterogeneities in temperature, composition, and phase are required to explain seismic observations. Several geodynamic hypotheses have arisen to explain these observations in the context of larger-scale mantle convection. Increased collaboration among disciplines involved in deep Earth interior research will be helpful in resolving the fundamental questions that remain.

References and Notes

1. T. Lay, D. V. Helmberger, Geophys. J. R. Astron. Soc. 75, 799 (1983).
2. R. D. van der Hilst, S. Widiyantoro, E. R. Engdahl, Nature 386, 578 (1997).
3. S. P. Grand, R. D. van der Hilst, S. Widiyantoro, GSA Today 7, 1 (1997).
4. J. Trampert, F. Deschamps, J. Resovsky, D. A. Yuen, Science 306, 853 (2004).
5. M. Ishii, J. Tromp, Phys. Earth Planet. Int. 146, 113 (2004).
6. S. Ni, E. Tan, M. Gurnis, D. V. Helmberger, Science 296, 1850 (2002).
7. Y. Wang, L. Wen, J. Geophys. Res. 112, B05313 (2007).
8. E. J. Garnero, T. Lay, A. K. McNamara, in The Origin of Melting Anomalies: Plates, Plumes and Planetary Processes, G. R. Foulger, D. M. Jurdy, Eds. (Geological Society of America, Boulder, CO, 2007), pp. 79–102.
9. A. K. McNamara, S. Zhong, Nature 437, 1136 (2005).
10. P. J. Tackley, Geochem. Geophys. Geosyst. 3, 10.1029/2001GC000167 (2002).
11. A. Davaille, E. Stutzmann, G. Silveira, J. Besse, V. Courtillot, Phys. Earth Planet. Sci. Lett. 239, 233 (2005).
12. A. M. Jellinek, M. Manga, Rev. Geophys. 42, RG3002 (2004).
13. E. Tan, M. Gurnis, Geophys. Res. Lett. 32, L20307 (2005).
14. E. Knittle, R. Jeanloz, Science 251, 1438 (1991).
15. U. R. Christensen, A. W. Hofmann, J. Geophys. Res. 99, 19867 (1994).
16. K. Hirose, N. Takafuji, N. Sata, Y. Ohishi, Earth Planet. Sci. Lett. 237, 239 (2005).
17. T. Lay, J. Hernlund, E. J. Garnero, M. S. Thorne, Science 314, 1272 (2006).
18. M. Thorne, E. J. Garnero, S. Grand, Phys. Earth Planet. Int. 146, 47 (2004).
19. R. Montelli et al., Science 303, 338 (2004); published online 4 December 2003 (10.1126/science.1092485).
20. R. D. van der Hilst, M. V. De Hoop, Geophys. J. Int. 167, 1211 (2006).
21. C. Thomas, E. J. Garnero, T. Lay, J. Geophys. Res. 109, 10.1029/2004JB003013 (2004).
22. A. Hutko, T. Lay, E. J. Garnero, J. S. Revenaugh, Nature 441, 333 (2006).
23. R. D. van der Hilst et al., Science 315, 1813 (2007).
24. M. Weber, J. P. Davis, Geophys. J. Int. 102, 231 (1990).
25. T. Kito, S. Rost, C. Thomas, E. J. Garnero, Geophys. J. Int. 169, 631 (2007).
26. M. Wysession et al., in The Core-Mantle Boundary Region, M. Gurnis, M. Wysession, E. Knittle, B. Buffett, Eds. (American Geophysical Union, Washington, DC, 1998), pp. 273–298.
27. M. Murakami, K. Hirose, K. Kawamura, N. Sata, Y. Ohishi, Science 304, 855 (2004); published online 8 April 2004 (10.1126/science.1095932).
28. A. R. Oganov, S. Ono, Nature 430, 445 (2004).
29. T. Tsuchiya, J. Tsuchiya, K. Umemoto, R. M. Wentzcovitch, Earth Planet. Sci. Lett. 224, 241 (2004).
30. J. W. Hernlund, C. Thomas, P. J. Tackley, Nature 434, 882 (2005).
31. T. Nakagawa, P. J. Tackley, Geophys. Res. Lett. 31, L16611 (2004).
32. J. Wookey, S. Stackhouse, J.-M. Kendall, J. Brodholt, G. D. Price, Nature 438, 1004 (2005).
33. T. Lay, E. J. Garnero, in Post-Perovskite: The Last Mantle Phase Transition, K. Hirose, D. Yuen, T. Lay, J. Brodholt, Eds. (American Geophysical Union, Washington, DC, 2007), pp. 129–154.
34. T. Lay et al., in The Core-Mantle Boundary Region, M. Gurnis, M. Wysession, E. Knittle, B. Buffett, Eds. (American Geophysical Union, Washington, DC, 1998), pp. 299–318.
35. E. J. Garnero, M. M. Moore, T. Lay, M. J. Fouch, J. Geophys. Res. 109, 10.1029/2004JB003004 (2004).
36. S. Merkel et al., Science 316, 1729 (2007).
37. M. Wallace, C. Thomas, Phys. Earth Planet. Int. 151, 115 (2005).
38. M. S. Thorne, E. J. Garnero, J. Geophys. Res. 109, B08301 (2004).
39. Q. Williams, E. J. Garnero, Science 273, 1528 (1996).
40. S. Rost, E. J. Garnero, Q. Williams, M. Manga, Nature 435, 666 (2005).
41. B. A. Buffett, E. J. Garnero, R. Jeanloz, Science 290, 1338 (2000).
42. W. L. Mao et al., Science 312, 564 (2006).
43. J. Ritsema, H. J. van Heist, J. H. Woodhouse, J. Geophys. Res. 109, B02302 (2004).
44. J. Badro et al., Science 305, 383 (2004).
45. J.-F. Lin et al., Science 317, 1740 (2007).
46. We thank J. Ritsema for the tomographic model. Supported by NSF grants EAR-0711401 and EAR-0453944 (E.J.G.) and by NSF grants EAR-0510383 and EAR-0456356 (A.K.M.).

Fire-Derived Charcoal Causes Loss of Forest Humus

David A. Wardle,* Marie-Charlotte Nilsson, Olle Zackrisson

Boreal forests serve as important global sources or sinks of carbon (C), and wildfire is a major driver of C storage in these forests. Although fire releases CO_2 to the atmosphere, it also converts plant biomass into forms of black carbon, such as charcoal, that are resistant to microbial attack and persist in the soil for thousands of years (1). It has frequently been suggested that, because of its resistance, black C can serve as an important long-term C sink that may help offset the release of human-induced CO_2 to the atmosphere (2, 3). However, charcoal is not biologically inert and can have important effects on soil biological processes (4, 5). The influence of charcoal on the decomposition of native soil organic matter remains poorly understood.

We conducted a simple experiment in each of three contrasting boreal sites in northern Sweden (6). Mesh bags were filled with pure humus collected from the forest, pure charcoal created in the laboratory, or a 50:50 mixture of humus plus charcoal (6). These bags were left in the field and harvested over 10 years. This approach is conceptually identical to that used for the litter-mix studies that have greatly advanced our understanding of the consequences of mixing different litter types (7). This approach allows comparisons of observed values in the mixture with what would be expected on the basis of each of the components of the mixture considered separately.

We found that, over the 10-year period, loss of mass and C from the bags containing mixtures of charcoal and humus was substantially greater than what was expected on the basis of the components considered separately [Fig. 1, Mix (obs)

versus Mix (exp)]. Further, nitrogen immobilization was less than expected in the mixture bags (Fig. 1). For these measurements, substrate mixing effects [i.e., values for (observed – expected) never differed significantly across the three sites [P value always greater than 0.20 according to analysis of variance (ANOVA)]. This result is despite the sites differing in both stand history and soil fertility (6) and points to similar effects of charcoal across contrasting sites. Given that charcoal decomposition rates in soil are extremely slow (2, 8) and that in our study system charcoal persists in the humus layer without evidence of mass loss (4), most of the enhanced loss of mass and C caused by mixing charcoal and humus must have resulted from charcoal promoting humus loss rather than humus promoting charcoal loss.

Substrate (i.e., glucose)–induced respiration (SIR), a relative measure of active microbial biomass (6), was always significantly greater in the mixture bags than the value predicted on the basis of the charcoal and humus components considered separately [Fig. 1, Mix (obs) versus Mix (exp)]. These results are consistent with charcoal particles serving as foci for adsorption of organic compounds and microbial growth and activity (4, 5), leading to enhanced decomposition rates and mass loss of associated humus. The enhanced microbial activity in the mixture bags may have led to greater mass and C loss through either greater respiration or greater leaching of soluble compounds (9).

Previous short-term laboratory studies have shown that charred plant material causes accelerated breakdown of simple carbohydrates (10).

Our results extend these findings by indicating that charcoal can promote rapid loss of forest humus and belowground C during the first decade after its formation. Fire often causes substantial losses of ecosystem C, and our results provide evidence for a previously unreported mechanism that could contribute to these losses. Our results are specific to boreal forests and to the type of charcoal that we used, and further work is needed to determine the importance of this mechanism in other biomes and for other types of charcoal (11). Although several studies have recognized the potential of black C for enhancing ecosystem C sequestration (2, 3), our results show that these effects can be partially offset by its capacity to stimulate loss of native soil C, at least for boreal forests. The effect of charcoal on native soil C needs to be explicitly considered to better understand the potential of black C as an ecosystem C sink and agent of C sequestration.

References and Notes

1. J. W. Harden et al., Global Change Biol. 6 (suppl.), 174 (2000).
2. M. W. Schmidt, A. C. Nowak, Global Biogeochem. Cycles 14, 777 (2000).
3. J. Lehmann, Nature 447, 143 (2007).
4. O. Zackrisson, M.-C. Nilsson, D. A. Wardle, Oikos 77, 10 (1996).
5. J. Pietikäinen, O. Kilkila, H. Fritze, Oikos 89, 231 (2000).
6. Materials and methods are available on Science Online.
7. B. Gartner, Z. G. Cardon, Oikos 104, 230 (2004).
8. C. M. Preston, M. W. Schmidt, Biogeoscience 3, 397 (2006).
9. J. C. Neff, D. U. Hooper, Global Change Biol. 8, 872 (2002).
10. U. Hamer, B. Marschner, S. Brodowski, Org. Geochem. 35, 823 (2004).
11. K. Hammes et al., Global Biogeochem. Cycles 21, GB3016 (2007).
12. We thank A. Sundberg, K. Boot, G. Rattray, and A. Mahamood for technical assistance and T. Fukami, M. Gundale, and anonymous reviewers for helpful comments.

Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, SE901-83 Umeå, Sweden.

*To whom correspondence should be addressed. E-mail: david.wardle@svek.slu.se
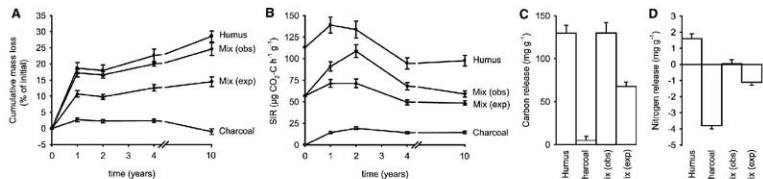
Fig. 1. Changes in litterbag properties over a 10-year period. Humus, Charcoal, and Mix (obs) correspond to litterbags containing pure humus, pure charcoal, and a 50:50 mixture of charcoal and humus, respectively. Mix (exp) corresponds to expected values for litterbags containing 50:50 mixtures of charcoal and humus if no interactive effects between the components occur (6). Each data point is the average of all three sites with 11 replicates per site, and vertical bars are mean within-site standard error. For all measurements at all dates and sites, values for Mix (exp) and Mix (obs) differ significantly at $P = 0.01$ (paired t tests). (**A**) Total mass loss. (**B**) SIR. (**C** and **D**) Loss of C and N from litter bags (per unit initial mass) over 10 years; negative values in (D) reflect net N gain through immobilization.

A Specialized Forebrain Circuit for Vocal Babbling in the Juvenile Songbird

Dmitriy Aronov, Aaron S. Andalman, Michale S. Fee*

Young animals engage in variable exploratory behaviors essential for the development of neural circuitry and adult motor control, yet the neural basis of these behaviors is largely unknown. Juvenile songbirds produce subsong—a succession of primitive vocalizations akin to human babbling. We found that subsong production in zebra finches does not require HVC (high vocal center), a key premotor area for singing in adult birds, but does require LMAN (lateral magnocellular nucleus of the nidopallium), a forebrain nucleus involved in learning but not in adult singing. During babbling, neurons in LMAN exhibited premotor correlations to vocal output on a fast time scale. Thus, juvenile singing is driven by a circuit distinct from that which produces the adult behavior—a separation possibly general to other developing motor systems.

How does a young brain learn to use the muscles it controls and the sensory organs by which it perceives the world? To a surprising extent, this knowledge is not built in by deterministic developmental rules but must be obtained through exploration. For instance, the relationship between feedback from the somatosensory periphery and movement is revealed to the developing brain by spontaneous muscle twitches, which facilitate the self-organization of spinal reflex circuits (1) and cortical somatosensory maps (2, 3). At a higher level, juvenile animals learn the causal relation between actions and the effects of these actions by producing highly variable behaviors such as infant movements, grasp-like "hand babbling," early vocalizations, and play (4–8).

How are these exploratory juvenile behaviors generated? Are they produced by the same brain areas responsible for the corresponding adult behaviors later in life, or are specialized brain regions involved? Forebrain areas, including the motor cortex and the basal ganglia, have been implicated in the production of normal infant movements, as well as their abnormalities (5, 9–11). Yet the specific forebrain circuits for infant motor control remain to be identified.

Babbling is an early motor behavior produced by juveniles of vocal mammals and birds (6, 12–15). In zebra finches, babbling, called subsong, occurs roughly from ages 30 to 45 days post-hatch (dph). Plastic song follows, with the gradual appearance of distinct and identifiable, but variable, vocal elements (syllables). By 80 to 90 dph, plastic song is gradually transformed into highly complex, stereotyped motifs—sequences

McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: fee@mit.edu

of syllables that constitute adult song. The premotor circuit for adult song production consists of HVC (high vocal center), RA (robust nucleus of the arcopallium), and brainstem motor nuclei (Fig. 1A). This "motor pathway" is crucial for generating stereotyped, learned vocalizations (16, 17) and exhibits firing that is precisely time-locked to the song output (18–21).

Another circuit, the anterior forebrain pathway (AFP), is homologous to basal ganglia thalamocortical loops in mammals and projects to RA through a forebrain nucleus, LMAN (lateral magnocellular nucleus of the nidopallium) (22, 23). Although LMAN is not required for singing in adult birds, it is necessary for normal song learning in juveniles (24, 25) and plays a role in producing song variability in adult and juvenile birds (26, 27). These and other studies have suggested a view that the motor pathway drives singing, whereas the output of the AFP modulates or instructs the motor pathway during learning (28, 29).

Subsong persists in the absence of HVC. We investigated whether primitive subsong vocalizations result from an immature form of the adult motor pathway, or whether they are driven by other premotor circuits. Given the importance of HVC for mature singing (16, 20, 30), we sought to characterize its involvement early in development. In nine subsong-producing juvenile birds (ages 33 to 44 dph), we eliminated HVC bilaterally, either by electrolytic lesions or by pharmacological inactivation (31). In three additional birds, we left HVC intact but specifically eliminated its projection to RA by bilateral transection of the HVC-to-RA fiber tract. After these manipulations, all birds continued producing largely unaffected subsong (Fig. 1A and fig. S3).

Surprisingly, older birds—those in the plastic-song stage (45 to 73 dph, n = 12) and adults (n = 5, undirected singing)—also sang after bilateral

HVC elimination [but see (31)]. These birds lost structure and stereotypy in their songs, reverting to the production of subsong-like vocalizations. After pharmacological inactivation of HVC, this reversion to subsong-like vocalizations was fast (within 20 min) and reversible (fig. S4); this finding suggested that the effect is not due to long-term changes in neural circuitry, but rather occurs immediately as a result of the loss of spiking activity in HVC. At all ages, singing in the absence of HVC was produced at normal rates and followed an ordinary circadian rhythm, with more songs produced in the morning than in later parts of the day (31).

Singing without HVC is highly similar to normal subsong. We asked whether the sounds produced in the absence of HVC were indeed similar to subsong. We characterized acoustic properties of songs by measuring spectral features shown to be effective for quantifying developmental trends in zebra finches (32, 33). Distributions of these features before and after HVC elimination were highly similar for subsong-producing birds (31). An additional feature of normal subsong is the absence of repeatable acoustic elements of stereotyped length. This was evident in a wide, unimodal distribution of syllable durations for subsong-producing birds (n = 9 birds younger than 45 dph; Fig. 2, A and B). After HVC elimination, these distributions were unchanged (31). In contrast, plastic and adult songs contain distinct syllables that form multiple narrow peaks in the distributions of durations. After HVC elimination in older birds, all distinct syllables were lost, resulting in unimodal distributions similar to those of subsong (n = 25 birds) (31).

Furthermore, subsong is characterized by a lack of sequential stereotypy, which appears later in plastic and adult songs. We quantified stereotypy by measuring the peak of the spectral cross-correlation between different song renditions (Fig. 2C) (31). In control conditions, stereotypy was higher for older birds (Fig. 2D; P < 0.0001 for nonzero slope of the linear regression of stereotypy and age). However, independently of age, stereotypy was reduced to the level of subsong after HVC elimination (Wilcoxon P > 0.1 for the difference from normal subsong). In summary, analyses of acoustic structure indicate that, by a wide range of measurements, singing in the absence of HVC is highly similar to normal subsong.

Subsong requires activity in RA and LMAN. If subsong persists in the absence of HVC, what neural circuits are engaged in its production? One possibility is that subsong does not require the forebrain song system and is entirely produced by midbrain or brainstem circuitry, even in the absence of RA. A second possibility is that subsong is driven by circuitry intrinsic to RA, even in the absence of HVC and LMAN. The third possibility is that subsong is driven by, or requires, inputs from LMAN to RA. We tested

these hypotheses by lesions and inactivations of RA and LMAN.

RA lesions entirely blocked singing in juvenile birds ($n = 5$, 39 to 73 dph), indicating that subsong-like vocalizations require descending inputs from the forebrain (Fig. 3). Similarly, song production was abolished by lesions of HVC and subsequent inactivation of LMAN ($n = 12$ experiments in 5 birds, 51 to 75 dph), indicating that RA circuitry, without its afferent inputs, is not sufficient to generate singing. We further tested the necessity of LMAN inputs to RA by inactivating LMAN in juvenile birds. LMAN inactivation entirely blocked subsong production in all birds younger than 45 dph ($n = 6$ experiments in 4 birds). However, in agreement with previous studies, LMAN inactivation did not block singing in most older birds (6 of 7 experiments in 5 birds, 45 to 67 dph), although it produced a marked reduction in song variability (26, 27). Together, these results indicate that RA and its inputs from LMAN are necessary for subsong production.

LMAN neurons exhibit premotor activity during subsong. An intriguing possibility suggested by the above results is that LMAN drives

subsong production—i.e., that it generates patterns of spiking activity that control the acoustic structure of subsong on a short (10 ms) time scale. To test this prediction directly, we recorded from single RA-projecting LMAN neurons during subsong production in intact birds [$n = 15$ neurons in 3 birds, 38 to 45 dph (31)] and in birds with bilateral HVC lesions ($n = 16$ neurons in 2 birds, lesioned at 38 and 50 dph). To quantify premotor activity, we examined firing in a short window preceding each syllable boundary (onset or offset). To begin with, we only considered syllable boundaries separated from other onsets or offsets by relatively long (>150 ms) periods to eliminate the possible confounding effects of neighboring syllables on the firing pattern. There was a significant increase in firing before syllable onsets in 12 of 31 neurons [16.1 ± 1.6 Hz in a 50-ms window preceding syllable onset versus 8.6 ± 0.6 Hz in a 100-ms baseline period preceding this window; $P < 0.05$; e.g., neuron 3, Fig. 4, A and B (31)]. Similarly, syllable offsets were preceded by a significant increase in firing in 5 of 31 neurons (21.2 ± 3.4 Hz before syllable offset versus baseline, $15.5 \pm$

1.3 Hz; $P < 0.05$; e.g., neuron 14, Fig. 4, C and D). Similar neuronal firing patterns related to onsets and offsets of behavioral sequences have been observed in other basal ganglia–related circuits (34).

In the above analysis, we only considered syllable boundaries separated by long (>150 ms) periods of time to isolate syllable onset- and offset-related changes in firing. However, the firing of some LMAN neurons also correlated with more rapid changes in song structure. For instance, neuron 12 (Fig. 4, E to G) exhibited increased firing before syllables that followed short (10 to 150 ms) rather than long intervals, as well as a reduction in firing during silent periods between syllables. Overall, seven neurons showed a premotor increase in activity before syllables separated by short intervals ($P < 0.05$ for the comparison of a 30-ms window preceding a syllable with 30 ms of baseline). This finding suggests that some LMAN neurons may have a premotor relation to subsong structure at the level of individual syllables.

In neurons that exhibited a significant increase in firing before syllable onsets ($n = 18$),
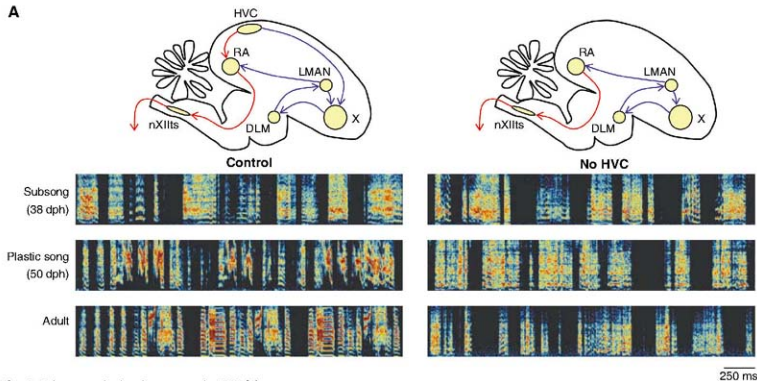
Fig. 1. Subsong production does not require HVC. (**A**) Results of bilateral HVC elimination (by lesion or pharmacological inactivation). Top: major connections of the song system with and without HVC. Red, motor pathway; blue, anterior forebrain pathway (AFP); X, area X, a basal-ganglia homolog; DLM, dorsolateral nucleus of the anterior thalamus; nXIIts, tracheosyringeal portion of the hypoglossal nucleus. Lower left: Sonograms of three birds at different ages. Lower right: Sonograms of the same birds in the absence of HVC. Frequency ranges from 500 Hz to 7.5 kHz; color scale (from black to red) spans a power range of 8 dB. For audio clips of these songs, see (31). (**B**) Histological verification of HVC lesions. Left: Inverted dark-field image of a parasagittal section of a normal zebra finch brain (50 dph). Red indicates retrograde fluorescence labeling of neurons in HVC.

after tracer (Alexa-conjugated cholera toxin subunit β) injection into RA. Inset: retrograde labeling of neurons in LMAN from the same injection. Right: Brain sections of the plastic-song bird shown in (A). Scale bars, 500 μm.
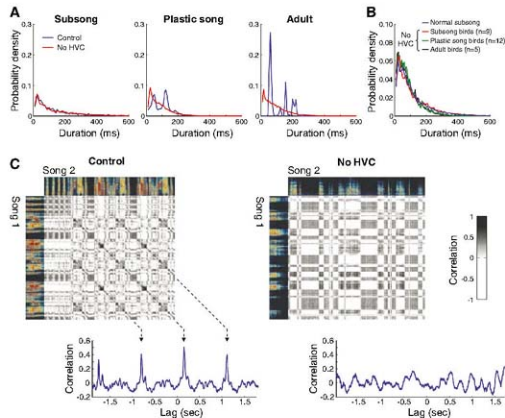
Fig. 2. Singing in the absence of HVC is highly similar to normal subsong. (**A**) Distributions of syllable durations for three birds of various ages (blue) and distributions for the same birds in the absence of HVC (red). (**B**) Average syllable duration distributions for normal subsong-producing birds (blue) and birds of different ages in the absence of HVC. (**C**) Sample spectral correlation matrices for a pair of songs produced by an adult bird (left) and by the same bird after HVC lesion (right). Averaging the matrix along its diagonals reveals strong correlation peaks in control (pre-lesion) condition, but not after HVC lesion. (**D**) Maximum values of the spectral correlation, averaged across all pairwise comparisons of 10 song bouts (31), for birds in control conditions and for the same birds in the absence of HVC. Dashed lines, linear regression; error bars, SEs across all 45 pairwise comparisons.

Fig. 3. Subsong production requires LMAN and RA. Average rates of song and call production in all lesion and inactivation experiments are shown. For rate measurement, a full day of recording was partitioned into 1-s segments, and the numbers of segments containing calls or songs were estimated (31). In cases where age is unspecified, data from all birds are pooled together. Note that for subsong-producing birds (<45 dph), the average rate of singing was not affected by HVC elimination (Wilcoxon P > 0.5). LMAN lesions in older juveniles (rightmost group) resulted in highly stereotyped song (27). Values at top are fractions of experiments in which any amount of singing occurred. Error bars are SEM values across birds. In experiments that abolished singing, silencing was specific to songs and did not affect the frequency of call vocalizations that are known not to require the song system (17).
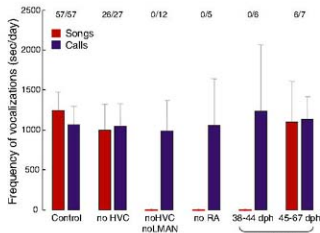
high-frequency bursts of spikes (>100 Hz) preceded 13.2 ± 1.4% of syllables. The most likely timing of a burst onset was 17.2 ± 3.1 ms before syllable onset. Such latency is, in fact, anticipated for premotor activity in LMAN, given the 10- to 15-ms latency reported for vocal perturbation after electrical stimulation in RA (35) and the 2- to 5-ms antidromic latency from RA we found in LMAN neurons (31). Note that although the exact relationship of firing to song varied across cells, 20 of 31 neurons we recorded (65%) showed some type of premotor correlation to the vocal output. Premotor firing in LMAN did not require activity within HVC; 8 of 16 neurons exhibited significant correlations to song structure in HVC-lesioned birds (fig. S5).

Discussion. Our data indicate that LMAN, and possibly other components of the AFP, constitute an essential premotor circuit for the production of early babbling. At the same time, we have shown that the classical premotor nucleus HVC (16) is not necessary for the generation of subsong. We therefore propose that two premotor pathways in the songbird function to produce vocalizations at different stages of development. In young juveniles, the AFP generates poorly structured subsong, whereas in adult birds, the classical HVC-motor pathway generates highly stereotyped motor sequences. These pathways interact in the intermediate plastic-song stage (27) to generate the partially structured but variable vocalizations upon which vocal learning operates.

The transfer of functional dominance from one pathway to another during vocal learning elegantly parallels their anatomical development. HVC does not reach its adult size until the late plastic-song stage (36) and establishes functional synapses in RA later than LMAN does (37, 38). Song maturation and the decrease in vocal variability have thus been attributed to the strengthening of inputs from HVC and the concurrent weakening of inputs from LMAN (39–42). Curiously, although HVC neurons form synapses in RA around the onset of singing [30 to 35 dph (37)], our results show that they do not significantly contribute to song production in its earliest stage. It is therefore possible that the HVC-to-RA pathway is active during early subsong but is not yet functionally strong enough to drive singing by itself or to influence vocalizations in a detectable way.

Identifying forebrain circuits involved in the production of juvenile behaviors is a requisite step toward understanding the mechanisms by which sensorimotor learning takes place. Several models of developmental learning suggest that early motor behaviors originate in the same circuits that later produce adult behavior. In this view, known as neuronal group selection theory, an initially large number of motor patterns undergo a selection process through competition, gradually eliminating circuits that produce undesirable behaviors (9, 43–46). Our findings, however, suggest a rather different model in which distinct specialized circuits are dedicated to the

Fig. 4. LMAN exhibits premotor activity during subsong. (**A**) Activity of an RA-projecting LMAN neuron during subsong production. Blue segments indicate individual syllables. Instantaneous firing rate exhibits peaks before syllable onsets. (**B**) Examples of spiking activity (red) before onset of sound amplitude (black) for neuron 3. Asterisk indicates a matching example with (A). Histograms show average firing rate across all syllable onsets for neuron 3; blue trace, average sound amplitude. Average includes only those syllables that were preceded by long (>150 ms) periods of silence. (**C** and **D**) Activity of a neuron that exhibited peaks in firing before syllable offsets, plotted as in (A) and (B). Averages in (D) include only long (>150 ms) syllables that were followed by long (>150 ms) periods of silence in order to isolate offset-related changes in firing from onset-related changes. (**E**) Activity of a neuron that exhibited firing before syllable onsets after short (<150 ms) intervals, plotted as in (A). Bottom: Spiking activity (red) occurring before syllable onsets for neuron 12. (**F**) Averages of firing rate and sound amplitude for neuron 12, separately for syllables that followed short (10 to 150 ms) and long (>150 ms) intervals, plotted as in (B). (**G**) Syllable onset–centered spike raster for neuron 12. Raster is sorted according to the length of the interval that preceded the syllables; dashed lines indicate interval boundaries. Blue marks, spikes that occurred in high-frequency (>100 Hz) bursts; gray marks, spikes that occurred outside of bursts.
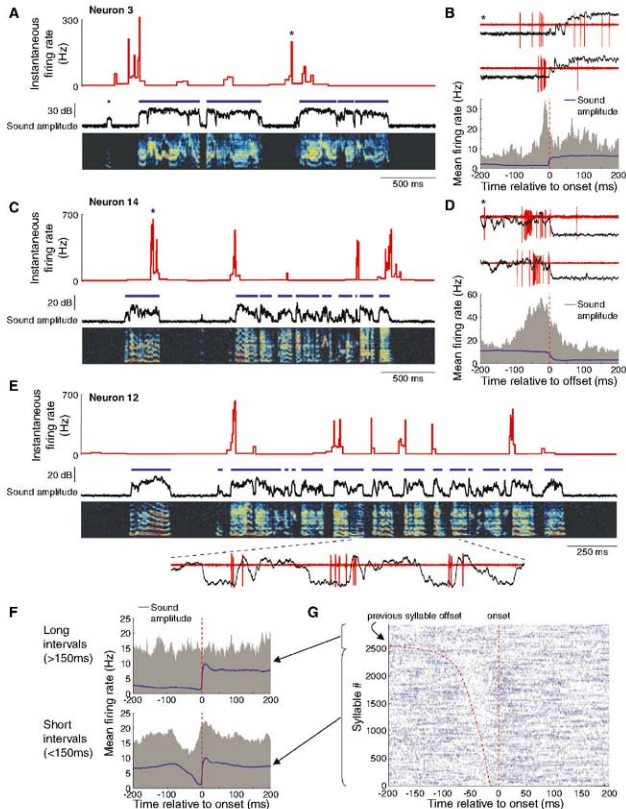
generation of highly variable juvenile behavior. We speculate that similar circuits for the production of infant behavior may be a general feature of developmental learning in the vertebrate brain.

References and Notes
1. P. Petersson, A. Waldenström, C. Fahraeus, J. Schouenborg, *Nature* **424**, 72 (2003).
2. R. Khazipov et al., *Nature* **432**, 758 (2004).
3. M. Milh et al., *Cereb. Cortex* **17**, 1582 (2007).
4. P. S. Wallace, I. Q. Whishaw, *Neuropsychologia* **41**, 1912 (2003).
5. S. R. Robinson, M. S. Blumberg, M. S. Lane, L. A. Kreber, *Behav. Neurosci.* **114**, 328 (2000).
6. A. J. Doupe, P. K. Kuhl, *Annu. Rev. Neurosci.* **22**, 567 (1999).
7. T. Imada et al., *Neuroreport* **17**, 957 (2006).
8. R. Fagen, *Animal Play Behavior* (Oxford Univ. Press, New York, 1981).
9. H. Forssberg, *Curr. Opin. Neurobiol.* **9**, 676 (1999).
10. J. A. Eyre, S. Miller, G. J. Clowry, E. A. Conway, C. Watts, *Brain* **123**, 51 (2000).
11. H. F. Prechtl et al., *Lancet* **349**, 1361 (1997).
12. P. Marler, *Am. Sci.* **58**, 669 (1970).
13. A. M. Elowson, C. T. Snowdon, C. Lazaro-Perea, *Trends Cogn. Sci.* **2**, 31 (1998).
14. M. Knornschild, O. Behr, O. von Helversen, *Naturwissenschaften* **93**, 451 (2006).
15. D. Reiss, B. McCowan, *J. Comp. Psychol.* **107**, 301 (1993).
16. F. Nottebohm, T. M. Stokes, C. M. Leonard, *J. Comp. Neurol.* **165**, 457 (1976).
17. H. B. Simpson, D. S. Vicario, *J. Neurosci.* **10**, 1541 (1990).

18. A. C. Yu, D. Margoliash, *Science* **273**, 1871 (1996).
19. Z. Chi, D. Margoliash, *Neuron* **32**, 899 (2001).
20. R. H. Hahnloser, A. A. Kozhevnikov, M. S. Fee, *Nature* **419**, 65 (2002).
21. A. Leonardo, M. S. Fee, *J. Neurosci.* **25**, 652 (2005).
22. S. W. Bottjer, K. A. Halsema, S. A. Brown, E. A. Miesner, *J. Comp. Neural.* **279**, 312 (1989).
23. M. A. Farries, D. J. Perkel, *J. Neurosci.* **22**, 3776 (2002).
24. S. W. Bottjer, E. A. Miesner, A. P. Arnold, *Science* **224**, 901 (1984).
25. C. Scharff, F. Nottebohm, *J. Neurosci.* **11**, 2896 (1991).
26. M. H. Kao, A. J. Doupe, M. S. Brainard, *Nature* **433**, 638 (2005).
27. B. P. Ölveczky, A. S. Andalman, M. S. Fee, *PLoS Biol.* **3**, e153 (2005).
28. K. Doya, T. J. Sejnowski, in *Advances in Neural Information Processing Systems*, G. Tesauro, D. S. Touretzky, T. K. Leen, Eds. (MIT Press, Cambridge, MA, 1995), vol. 7, pp. 101–108.

29. T. W. Troyer, S. W. Bottjer, *Curr. Opin. Neurobiol.* **11**, 721 (2001).
30. J. A. Thompson, F. Johnson, *Dev. Neurobiol.* **67**, 205 (2007).
31. See supporting material on *Science* Online.
32. O. Tchernichovski, F. Nottebohm, C. E. Ho, B. Pesaran, P. P. Mitra, *Anim. Behav.* **59**, 1167 (2000).
33. S. Deregnaucourt, P. P. Mitra, O. Feher, C. Pytte, O. Tchernichovski, *Nature* **433**, 710 (2005).
34. N. Fujii, A. M. Graybiel, *Science* **301**, 1246 (2003).
35. M. S. Fee, A. A. Kozhevnikov, R. H. Hahnloser, *Ann. N.Y. Acad. Sci.* **1016**, 153 (2004).
36. A. Alvarez-Buylla, C. Y. Ling, F. Nottebohm, *J. Neurobiol.* **23**, 396 (1992).
37. R. Mooney, *J. Neurosci.* **12**, 2464 (1992).
38. R. Mooney, M. Rao, *J. Neurosci.* **14**, 6532 (1994).
39. K. Herrmann, A. P. Arnold, *J. Neurosci.* **11**, 2063 (1991).
40. E. Akutagawa, M. Konishi, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12413 (1994).
41. J. M. Kittelberger, R. Mooney, *J. Neurosci.* **19**, 9385 (1999).
42. L. L. Stark, D. J. Perkel, *J. Neurosci.* **19**, 9107 (1999).

43. G. M. Edelman, *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic Books, New York, 1987).
44. O. Sporns, G. M. Edelman, *Child Dev.* **64**, 960 (1993).
45. P. Marler, *J. Neurosci. Methods* **33**, 501 (1997).
46. M. Hadders-Algra, *Dev. Med. Child Neurol.* **42**, 566 (2000).
47. We thank A. Graybiel, E. Bizzi, and J. Goldberg for comments on the manuscript and F. Nottebohm for helpful discussion regarding HVC lesions. Supported by NIH grant MH067105, a Hertz Foundation Silvio Micali fellowship (D.A.), and a Friends of the McGovern Institute fellowship (A.S.A.).

Supporting Online Material
www.sciencemag.org/content/full/320/5876/630/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Sounds S1 to S9
References

11 January 2008; accepted 24 March 2008
10.1126/science.1155140

High-Thermoelectric Performance of Nanostructured Bismuth Antimony Telluride Bulk Alloys

Bed Poudel,[1,2]* Qing Hao,[3]* Yi Ma,[1,2] Yucheng Lan,[1] Austin Minnich,[3] Bo Yu,[1] Xiao Yan,[1] Dezhi Wang,[1] Andrew Muto,[3] Daryoosh Vashaee,[3] Xiaoyuan Chen,[3] Junming Liu,[4] Mildred S. Dresselhaus,[5] Gang Chen,[3]† Zhifeng Ren[1]†

The dimensionless thermoelectric figure of merit (ZT) in bismuth antimony telluride (BiSbTe) bulk alloys has remained around 1 for more than 50 years. We show that a peak ZT of 1.4 at 100°C can be achieved in a p-type nanocrystalline BiSbTe bulk alloy. These nanocrystalline bulk materials were made by hot pressing nanopowders that were ball-milled from crystalline ingots under inert conditions. Electrical transport measurements, coupled with microstructure studies and modeling, show that the ZT improvement is the result of low thermal conductivity caused by the increased phonon scattering by grain boundaries and defects. More importantly, ZT is about 1.2 at room temperature and 0.8 at 250°C, which makes these materials useful for cooling and power generation. Cooling devices that use these materials have produced high-temperature differences of 86°, 106°, and 119°C with hot-side temperatures set at 50°, 100°, and 150°C, respectively. This discovery sets the stage for use of a new nanocomposite approach in developing high-performance low-cost thermoelectric materials.

S olid-state cooling and power generation based on thermoelectric effects have potential applications in waste-heat recovery, air conditioning, and refrigeration. The efficiency of thermoelectric devices is determined by the materials' dimensionless figure of merit, defined as $ZT = (S^2\sigma/k)T$, where S, σ, k, and T are the Seebeck coefficient, electrical conductivity, thermal conductivity, and absolute temperature, respectively (1–3). An average ZT in the

application temperature range must be higher than 1 to make a thermoelectric device competitive (1–3).

There have been persistent efforts to improve ZT values since the 1950s, but the peak ZT of dominant commercial materials based on Bi_2Te_3 and its alloys, such as $Bi_xSb_{2-x}Te_3$ (p-type), has remained at 1. During the past decade, several groups have reported enhanced ZT in (i) superlattices such as Bi_2Te_3/Sb_2Te_3 (4) and $PbSe_{0.98}Te_{0.02}/PbTe$ (5), because of reductions in the lattice thermal conductivity, and (ii) new bulk materials, such as lead antimony silver telluride (LAST) and its alloys (6), including skutterudites (7). Although high ZT values were reported in superlattice structures, it has proven difficult to use them in large-scale energy-conversion applications because of limitations in both heat transfer and cost. Bulk materials with improved ZT, such as LAST and skutterudites, are ideal for high-temperature operations. However, at rela-

tively near room temperature (0° to 250°C), Bi_2Te_3-based materials still dominate.

We have pursued an approach in which the primary cause of ZT enhancement in superlattices—reduced thermal conductivity—also exists in random nanostructures (8, 9). We report a substantial ZT increase in bulk materials made from nanocrystalline (NC) powders of p-type $Bi_xSb_{2-x}Te_3$, reaching a peak ZT of 1.4 at 100°C. The enhanced ZT is the result of a significant reduction in thermal conductivity caused by strong phonon scattering by interfaces in the nanostructures. There have also been reports of ZT improvements at room temperature in Bi_2Te_3-based materials caused by the addition of Bi_2Te_3 nanotubes (10) and by melt spinning (11).

Our method, on the other hand, is based on the ball milling and hot pressing of nanoparticles into bulk ingots. This approach is simple, is cost effective, and can be used on other materials. Our materials have a ZT of about 1.2 at room temperature and 0.8 at 250°C with a peak of 1.4 at 100°C. In comparison, conventional Bi_2Te_3-based materials have a peak ZT of about 1 at room temperature and about 0.25 at 250°C. The high ZT in the 25° to 250°C temperature range makes the NC bulk materials attractive for cooling and low-grade waste-heat recovery applications. The materials can also be integrated into segmented thermoelectric devices for thermoelectric power generation that operate at high temperatures. In addition to the high ZT values, the NC bulk materials are also isotropic. They do not suffer from the cleavage problem that is common in traditional zone melting–made ingots, which leads to easier device fabrication and system integration and to a potentially longer device lifetime.

Sample preparation. Nanopowders were made by ball milling bulk p-type BiSbTe alloy ingots (12). Bulk disk samples (1.25 to 2.5 cm in diameter and 2 to 15 mm in thickness) were made by hot pressing the nanopowders loaded in 1.25- to 2.5-cm (inner diameter) graphite dies (12). Disks (1.25 cm in diameter and 2 mm in thickness) and bars (about 2 mm by 2 mm by

[1]Department of Physics, Boston College, Chestnut Hill, MA 02467, USA. [2]GMZ Energy, Incorporated, 12A Hawthorn Street, Newton, MA 02458, USA. [3]Department of Mechanical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [4]Laboratory of Solid State Microstructures and Department of Physics, Nanjing University, China. [5]Department of Physics and Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: gchen2@mit.edu (G.C.); renzh@bc.edu (Z.R.)

12 mm) were cut from both the axial and disk plane directions. The disks and bars were also polished for σ, k, and S measurements.

To achieve high ZT, researchers must control the size and quality of the starting nanoparticles. For good electrical conductivity, it is especially

Fig. 1. XRD (**A**), SEM (**B**), low-magnification TEM (**C**), and high-magnification TEM (**D**) images of an as–ball-milled nanopowder. a.u., arbitrary units.

important to prevent oxidation. Figure 1 shows the x-ray diffraction (XRD) pattern (Fig. 1A), the scanning electron microscope (SEM) image (Fig. 1B), and the low- and high-magnification transmission electron microscope (TEM) images (Fig. 1, C and D) of the nanopowders after ball

milling. The XRD patterns verify that the powder is in a single phase and is well matched with $Bi_{0.5}Sb_{1.5}Te_3$. The broadened diffraction peaks indicate that the particles are small, which is also confirmed in the SEM image (Fig. 1B) and the low-magnification TEM image (Fig. 1C). The TEM image (Fig. 1C) also shows that the nanoparticles have sizes ranging up to 50 nm, with an average size of about 20 nm. The high-resolution TEM image (Fig. 1D) confirms the excellent crystallinity of the nanoparticles and the clean surfaces. The inset in Fig. 1D also shows that some of the nanoparticles are even smaller than 5 nm.

Transport properties. The temperature dependences of several key properties of a typical NC bulk sample are compared in Fig. 2 with those of the state-of-the-art (SOA) p-type BiSbTe alloy ingot. All of the properties were measured in the same direction and reproduced on about 100 samples. The electrical conductivity of the NC bulk sample is slightly higher than that of the SOA ingot (Fig. 2A), but the Seebeck coefficient of the bulk sample is either slightly lower or higher than that of the ingot, depending on its temperature (Fig. 2B). Ultimately, the power factor ($S^2\sigma$) values of the bulk sample are similar to or higher than those of an ingot at temperatures below 50°C and above 75°C, respectively (Fig. 2C).

We also found that the thermal conductivity of the NC bulk samples is significantly lower than that of the ingot and, more importantly, that the difference increases with increasing temperature (Fig. 2D), which leads to significantly enhanced ZT values (Fig. 2E) in the 20° to 250°C temperature range. It also shows that the peak ZT value shifts to a higher temperature (100°C). The peak ZT of the NC bulk samples is about
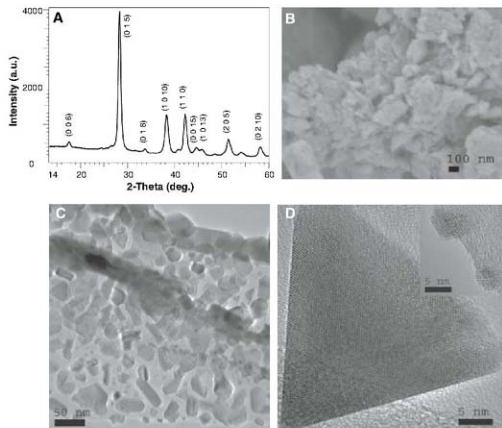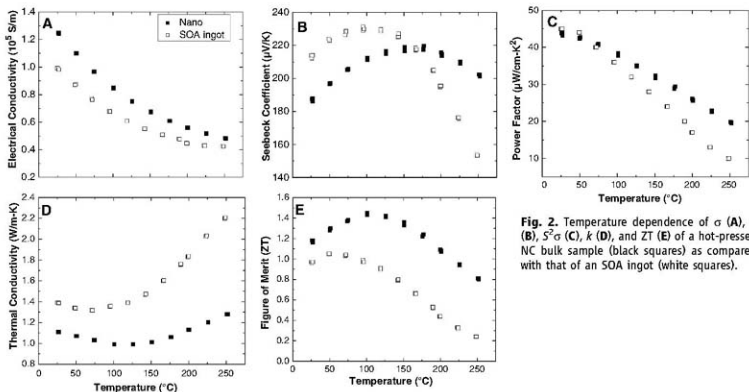
Fig. 2. Temperature dependence of σ (**A**), S (**B**), $S^2\sigma$ (**C**), k (**D**), and ZT (**E**) of a hot-pressed NC bulk sample (black squares) as compared with that of an SOA ingot (white squares).

1.4 at 100°C, which is significantly greater than that of the SOA Bi_2Te_3-based alloys. The ZT value of the SOA ingot starts to drop above 75°C and is below 0.25 at 250°C, whereas the ZT values for the NC bulk samples are still above 0.8 at 250°C. Such ZT characteristics are suitable for power generation applications because of a lack of available materials with high ZT in this temperature range.

All of these measurements were confirmed by two independent techniques on more than 100 samples. The electrical conductivity was measured by a four-point current-switching technique. We measured the Seebeck coefficient by a static dc method based on the slope of a voltage versus temperature-difference curve, using commercial equipment (ZEM-3, Ulvac, Inc., Methuen, Massachusetts, USA) on the same bar-type sample with a cross-sectional dimension of 2 mm by 2 mm and a length of 12 mm. The properties in the same sample were also measured by a home-built system, and the two sets of measurements are within 5% of each other.

We first measured the thermal diffusivity α by a laser-flash method on a disk using a commercial system (Netzsch Instruments, Inc., Burlington, Massachusetts, USA). After the measurement, bars were diced from the disks and α values were measured via the Ångstrom method in the

same home-built system. The α values from the bar and the disk agree within a range of 5%.

The thermal conductivity was calculated via the equation $k = \alpha \rho c_p$, where ρ is the density and c_p is the specific heat of the material that was measured with a differential scanning calorimeter (Netzsch Instruments, Inc.). To further check the property isotropy of the NC bulk samples, we cut disks and bars along and perpendicular to the press direction and then performed the measurements. Although individual properties may differ by 5% within the two directions, the final ZT values are isotropic. Such nearly isotropic characteristics are the result of the random orientation of the nanograins, showing that our NC bulk materials are superior to zone melting–made SOA Bi_2Te_3-based alloys, which have layered structures and, consequently, anisotropic thermoelectric properties. The highest ZT of the zone melting–made materials is along the basal plane of the crystal, but unfortunately this direction is vulnerable to cleavage, resulting in a difficult and low-yield module fabrication process.

For cooling applications, the stability of the nanostructures should not be a serious concern. High-temperature stability testing on the current NC bulk materials by repeated measurements up to 250°C did not show any sign of degradation,

suggesting potential power generation applications for waste-heat recovery in the reported temperature range.

Microstructure of NC bulk ingots. Detailed microstructure studies by TEM were carried out on NC bulk samples. The TEM specimens were prepared by dicing, polishing, and ion milling the bulk samples (13). Figure 3 shows the main structural features that we observed. In general, most of the grains are nanosized (Fig. 3, A and B). Furthermore, these nanograins are highly crystalline, are completely random (large angles between adjacent lattice planes), and have very clean boundaries between grains. They are also closely packed (Fig. 3B), which is consistent with our full density measurements. We also observed some larger grains (Fig. 3C). However, under high-resolution TEM observation, these grains consist of 2- to 10-nm-sized nanodots with fuzzy boundaries (Fig. 3D). Usually, these nanodots are Sb-rich with a typical composition close to that of Bi:Sb:Te = 8:44:48, with Sb substituted for Te.

Although some of the nanodots are without boundaries (Fig. 3D), we found other nanodots that make small-angle boundaries with the matrix (Fig. 3E). In addition, we also observed pure Te precipitates ranging from 5 to 30 nm in size (Fig. 3F). The selected-area electron diffraction (SAED) pattern (Fig. 3F, inset) confirms the presence of
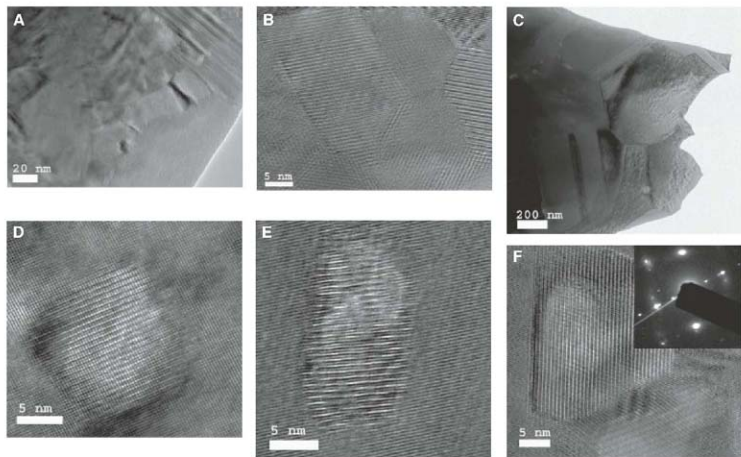
Fig. 3. TEM images showing the microstructures of a hot-pressed NC bulk sample. (**A**) Low-magnification image showing the nanograins. (**B**) High-magnification image showing the nanosize, high crystallinity, random orientation, and clean grain boundaries. (**C**) Low-magnification image showing larger grains. (**D** and **E**) High-magnification images showing the nanodots in the matrix without boundaries (D) and with small-angle grain boundaries (E). (**F**) High-magnification image showing Te precipitate in the matrix. SAED pattern (inset) shows the Te phase of the precipitated nanodot in the matrix.

this Te phase. Generally speaking, nanodots can be found within a 50-nm-diameter area.

We speculate that these nanodots could be formed during the hot-press heating and cooling processes. Similar types of nanodots have been observed in LAST alloys and were allegedly responsible for the ZT enhancement in those alloys (6). However, because there are so many interfaces from nanograins in our material, nanodots may not be the only reason for the strong phonon scattering. The larger-sized grains containing nanodots (Fig. 3C) are likely to be the result of the nonuniform ball milling of the ingot and may have experienced some grain growth during the hot-press compaction via Oswald ripening. More uniform nanograins produced during ball milling may retain their nanosize during the hot-press processing.

In comparing the transport properties of the NC bulk samples with those of the SOA ingot, it is important to note the relatively slow increase in k as a function of temperature for the NC bulk samples (Fig. 2D). This increase indicates a smaller bipolar contribution (2) to the conductivity by thermally generated electrons and holes in the NC bulk materials. We explain this reduced bipolar effect by assuming the existence of an interfacial potential that scatters more electrons than holes. Past studies in Bi_2Te_3-related materials suggested that structural defects, such as antisites (i.e., Bi atoms go to Te sites), serve as an important doping mechanism (14, 15). We anticipate that such antisites are more likely to occur at interfaces. Uncompensated recombination centers at interfaces associated with defect states and antisites are responsible for charge buildup at grain boundaries and thus increase the hole density in the grains. This explanation is consistent with the observed increase in the electrical conductivity as well as the reduction in the Seebeck coefficient of the NC bulk samples, as compared with those properties of the SOA ingot parent material (Fig. 2, A and B). We modeled the transport properties based

on the Boltzmann equation within the relaxation time approximation, including the interfacial potential, and we thus obtained the thermal conductivity contributions to the thermal conductivity shown in Fig. 4. The modeled results show that phonon contributions to the lattice thermal conductivity are reduced by a factor of two.

Cooling devices. To further confirm the ZT values, we constructed two unicouple cooling devices to measure their maximum temperature difference. One unicouple uses a nanostructured p-type material as one leg and a commercially available n-type material as the other leg (Fig. 5A). The performance of this unicouple was compared with that of a second unicouple made of commercially available n- and p-type materials from the same vendor. Testing was performed in a cryostat chamber with a typical pressure of 10^{-3} torr, so that the hot-side temperature could be controlled by a combination of a heater and a continuous

flow of liquid nitrogen. Figure 5B shows the temperature difference created at different current injection levels when the hot side is maintained at 100°C. The inset displays the temperature differences created by the two devices at different hot-side temperatures. The temperature differences created with the unicouple made of commercially available p-type material are consistent with values given in the vendor catalog, whereas the performance of the unicouple made of nanostructured p-type material is significantly better. Device modeling via the measured properties is consistent with our experimental results.

Conclusion. We have developed a nanostructuring approach, and we have demonstrated significant improvement in ZT in p-type BiSbTe NC bulk alloys. The value of ZT starts at 1.2 at room temperature, peaks to 1.4 at 100°C, and decreases to 0.8 at 250°C. Such ZT characteristics are attractive for both cooling and low-temperature

Fig. 4. Thermal conductivity of $Bi_xSb_{2-x}Te_3$ NC bulk alloys. White and black squares represent the experimental results for an SOA ingot and our NC bulk alloys, respectively. Solid lines represent the corresponding calculations of the total (top curves) and lattice (bottom curves) contribution to the thermal conductivity, respectively.
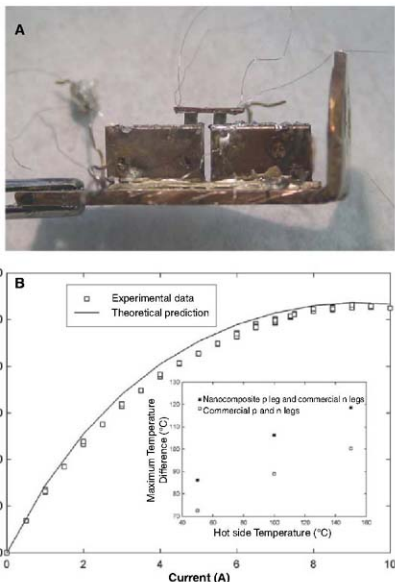
Fig. 5. (**A**) Experimental setup of the device cooling test. Two thermoelectric legs were mounted onto 6.5-mm-thick copper blocks and then bridged by a top copper plate. Two cold-side thermocouples were soldered into small holes drilled on both ends of the top copper plate. Hot-side thermocouples were soldered on the edge of each copper block. (**B**) Cooling test results with the hot side fixed at 100°C. The inset compares the maximum temperature difference of two unicouples with hot-side temperatures set at 50°, 100°, and 150°C.

waste-heat recovery applications. These materials are synthesized via ball milling, followed by dc hot pressing. The ZT enhancement comes mainly from a large reduction in the phonon thermal conductivity but also benefits from a reduction of bipolar contributions to the electronic thermal conduction at high temperatures. In the past, ZT enhancements have been reported in superlattice structures because of phonon thermal conductivity reduction. Our study suggests that a similar mechanism can be reproduced in random nanostructured bulk materials. Unlike superlattices and bulk crystals, a nanostructured BiSbTe alloy displays nearly isotropic ZT values. We further confirmed ZT enhancements by measuring the temperature difference created by unicouple devices constructed out of the materials for hotside temperature maintained between 50° to 150°C. The nanostructure synthesis method that we developed is a low-cost technique and can be readily scaled for mass production. These results provide a cost-effective means to improve the performance of thermoelectric materials.

References and Notes
1. D. M. Rowe, Ed. *CRC Handbook of Thermoelectrics* (CRC, Boca Raton, FL, 1995).
2. H. J. Goldsmid, *Thermoelectric Refrigeration* (Plenum, New York, 1964).
3. T. M. Tritt, Ed. *Semiconductors and Semimetals, Recent Trends in Thermoelectric Materials Research: Part One to Three* (Academic, San Diego, CA, 2001), vol. 69 to 71.
4. R. Venkatasubramanian, E. Silvola, T. Colpitts, B. O'Quinn, *Nature* 413, 597 (2001).
5. T. C. Harman, P. J. Taylor, M. P. Walsh, B. E. LaForge, *Science* 297, 2229 (2002).
6. K. F. Hsu *et al.*, *Science* 303, 818 (2004).
7. J. P. Fleurial, T. Caillat, A. Borshchevsky, in *Proceedings of the 13th International Conference on Thermoelectrics, Kansas City, MO, 30 August to 1 September 1994* (AIP, New York, 1995), pp. 40–44.
8. G. Chen, *Phys. Rev. B* 57, 14958 (1998).
9. M. S. Dresselhaus *et al.*, *Adv. Mater.* 19, 1043 (2007).
10. X. B. Zhao *et al.*, *Appl. Phys. Lett.* 86, 062111 (2005).
11. X. F. Tang, W. J. Xie, H. Li, W. Y. Zhao, Q. J. Zhang, *Appl. Phys. Lett.* 90, 012102 (2007).
12. Bulk p-type BiSbTe alloy ingots were loaded into a jar with balls inside the argon-filled glove box to avoid oxidation of the nanopowder. The jar was loaded into a ball mill and processed for several hours. When the nanopowder was ready, it was loaded into 1.25- to 2.5-cm (inner diameter) dies and compacted into a 100% dense solid NC bulk sample by a hot press. Samples are available for testing upon request.
13. We cut hot-pressed NC bulk pellets into blocks (2 mm by 3 mm by 1 mm) that were ground down into smaller blocks (2 mm by 3 mm by 0.002 mm) using a mechanical tripod polisher. We then glued the sample to a copper grid and milled it using a precision ion polishing system (Gatan Inc., Warrendale, Pennsylvania, USA) for 30 min, with incident energy of 3.2 kV and beam current of 15 µA at an incident angle of 3.5°.
14. T. C. Harman, S. E. Miller, H. L. Goering, *Bull. Am. Phys. Soc.* 30, 35 (1955).
15. T. Thonhauser, G. S. Jeon, G. D. Mahan, J. O. Sofo, *Phys. Rev. B* 68, 205207 (2003).
16. The work is supported by the U.S. Department of Energy (DOE) grant no. DE-FG02-00ER45805 (Z.F.R.), DOE grant no. DE-FG02-02ER45977 (G.C.), NSF–Nanoscale Interdisciplinary Research Team grant no. 0506830 (G.C., Z.F.R., and M.S.D.), National Science Foundation of China project no. 50528203 (J.M.L. and Z.F.R.), and the Ministry of Science and Technology of China project no. 2006CB921082 (J.M.L.). G.C. and Z.F.R. are cofounders of GMZ Energy, Inc.

13 February 2008; accepted 11 March 2008
Published online 20 March 2008;
10.1126/science.1156448
Include this information when citing this paper.

Coherent Control of Decoherence

Matthijs P. A. Branderhorst,[1] Pablo Londero,[1] Piotr Wasylczyk,[1] Constantin Brif,[2] Robert L. Kosut,[3] Herschel Rabitz,[2] Ian A. Walmsley[1]

Manipulation of quantum interference requires that the system under control remains coherent, avoiding (or at least postponing) the phase randomization that can ensue from coupling to an uncontrolled environment. We show that closed-loop coherent control can be used to mitigate the rate of quantum dephasing in a gas-phase ensemble of potassium dimers (K_2), which acts as a model system for testing the general concepts of controlling decoherence. Specifically, we adaptively shaped the light pulse used to prepare a vibrational wave packet in electronically excited K_2, with the amplitude of quantum beats in the fluorescence signal used as an easily measured surrogate for the purpose of optimizing coherence. The optimal pulse increased the beat amplitude from below the noise level to well above it, and thereby increased the coherence life time as compared with the beats produced by a transform-limited pulse. Closed-loop methods can thus effectively identify states that are robust against dephasing without any previous information about the system-environment interaction.

Interference is one of the hallmarks of quantum physics, and its presence is generally taken to demarcate the boundary between quantum and classical behavior. Controlling a quantum system consists of manipulating the relative amplitudes and phases of different distinct quantum states of the system, in order to achieve some objective, such as a particular functional operation or an increased yield of a particular outcome. The map from the optimal set of controls to the interference pattern of quantum probability amplitudes, which generates the desired outcome, is typically very complex. Therefore, iterative adaptive control (*1*), in which the system is incorporated into a feedback loop, is one of the most powerful tools for optimizing quantum control results in the laboratory. This method has wide currency and is applicable to a variety of very different physical and chemical systems and processes. For example, the closed-loop approach enabled the experimenters to determine the shape of an atomic electron's wave function (*2*), energy flow in a photosynthetic complex (*3*), polarization-sensitive photoionization channels (*4*), isotope-selective photoionization of molecules (*5*), photoisomerization of the retinal molecule in bacteriorhodopsin (*6*), high-harmonic generation of coherent soft x-rays (*7*), selective photodissociation and rearrangement of molecular bonds (*8*), and large-amplitude oscillations in C_{60} (*9*).

The effectiveness of coherent control is compromised by the coupling of the system to an uncontrolled environment, which disturbs in a random fashion the delicate quantum phases that define the state of the system. The time scales for decoherence vary widely. For example, the dephasing times for electronic dipole excitations range from the femtosecond to nanosecond regime, as a result of coupling to local phonon modes and spontaneous emission; those for spin (magnetic-dipole) excitations vary from microseconds to milliseconds, resulting from coupling to other spins in the sample or from collisions.

Typically, the quantum system is prepared in an initial state by the controller. The system then undergoes some free evolution, during which it is affected by the environment: Both loss of quantum coherence and dissipation of energy may occur. The inability to sustain the phase and amplitude relationships between the eigenstates of the system renders the controls ineffective. Any realistic goal for quantum control therefore requires sustaining the coherence in the system, and it is an open question whether and how this is possible in the face of dephasing perturbations. In particular, is there an optimal strategy to counter the effects of the environment? Protecting the quantum coherence is an important goal in various systems and processes. For example, it was recently discovered (*10*) that the coherence of excited electronic states plays an important role in biochemical processes, such as energy harvesting in photosynthesis. Similarly, the coherence of multiparticle superposition states is crucial for the implementation of quantum information processing (*11, 12*). It is in this context that the control of decoherence has been studied most com-

[1]Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK. [2]Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. [3]SCSolutions, 1261 Oakmead Parkway, Sunnyvale, CA 94085, USA.

*To whom correspondence should be addressed. E-mail: walmsley@physics.ox.ac.uk

prehensively, leading to the concepts of quantum error correction (QEC) (13, 14), dynamical decoupling (DD) (15, 16), and decoherence-free subspaces (DFSs) (17). These approaches either actively correct deviations of the system evolution from the one expected (QEC) or decouple the system from the environment, either actively (DD) or passively (DFS).

We wondered whether the tools of coherent control could themselves be used to inhibit dephasing. Because decoherence is ubiquitous, this is a far-reaching question, and the (perhaps) unexpected answer is that they can. We demonstrate experimentally some general principles that have broad application across many fields where maintaining quantum coherence is important. We chose a simple initial problem: maintaining the coherence of a prepared quantum state, without directing that coherence toward achievement of any particular function. This approach allowed us to explore experimentally some of the fundamental issues and to learn from the controls the mechanism by which they succeeded. The results give

confidence that the approach will be widely applicable. For example, recent theoretical studies suggest that optimal coherent control can help increase the purity of quantum states in dissipative atomic and molecular systems (18, 19), suppress dephasing due to electron-phonon scattering in semiconductor quantum dots (20, 21), reduce the rate of dissipation in a spin-boson model in the strong coupling regime (22), suppress various channels of electronic and vibrational dissipation and dephasing in molecules (23), and protect coherence in spin-based quantum gates (21, 24). An important conclusion of these theoretical studies, which is confirmed by our experimental results, is that the coherent preparation of quantum states can substantially alter the outcome of the subsequent nonunitary dynamics induced by the uncontrolled environment. Moreover, the method of closed-loop optimization successfully implemented in the present experiment has been proven to be a very efficient tool for finding optimal coherent controls in the laboratory. The generality of the closed-loop approach should make it useful

for management of coherence in a variety of problems (e.g., in those mentioned above), provided that easily measurable markers for coherence are available. These future applications depend on the progress in developing laboratory techniques for coherent manipulation and measurement of the corresponding quantum systems.

First, we wish to consider a system that can be analyzed completely and is experimentally tractable, yet is of sufficient complexity that its decoherence dynamics are essentially generic, so that the methods we use and results we obtain for this system are in principle applicable to a wide range of more complex situations. The criteria for identifying such a system are that there exists a set of feasible laboratory measurements that maps a particular degree of freedom of the system to experimental outcomes and that the unobserved part of the system has the known properties of a dephasing environment. We take the space of the corresponding set of observables to define the system itself. This essentially pragmatic approach is required because the system and environment are necessarily coupled; hence, there is no clear separation of the overall Hilbert space. A key criterion for the unobserved part of the entity (that is, the environment) is that it is of sufficient size that the entropy flow is essentially one way, with no environmentally induced revivals of the system coherence on the time scale of the experiment (25). Moreover, the spectrum of the environment must be broad and unstructured, thereby providing a short correlation time.

Second, a critical factor in the operation of a control loop is identifying an experimentally accessible, faithful measure of the objective characteristic, around which a control loop can be implemented. Standard measures of coherence—such as the purity of the state of the system, $P = \text{Tr}(\rho^2)$, or the von Neumann entropy, $S = -\text{Tr}(\rho \ln \rho)$—are nonlinear functions of the state operator ρ and can therefore be estimated only via joint measurements on pairs of identically prepared systems (23, 26) or calculated from state estimates determined by means of quantum state tomography (27). These approaches are technically challenging or may require too much data to be of use in a reasonable bandwidth feedback loop. Therefore, simple measures must be found that may act as a marker for coherence or a coherence surrogate (CS). That is, for a range of values of the CS around a certain threshold, coherence of the system is guaranteed and otherwise ruled out. One CS is the degree to which the system is localized or confined to a small region of configuration or phase space, because this property necessarily implies that the quantum state is a coherent superposition of the energy eigenstates of the system (28). A possible strategy to control decoherence is therefore to measure the localization of a wave packet at some time after its excitation and then to use this information in a feedback loop that changes the shape of the packet in such a manner as to maximize the longevity of its coherence. This
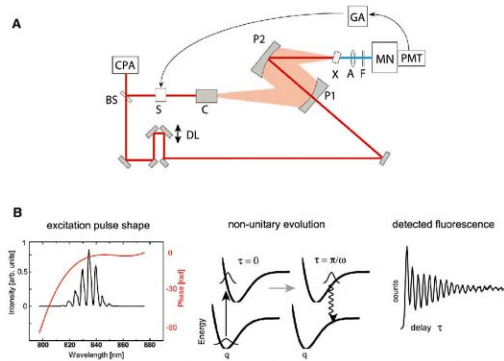
Fig. 1. (**A**) Apparatus for control and measurement of the vibrational wave packet states of potassium dimers and (**B**) excitation of a shaped vibrational wave packet and its detection at the outer turning point. A pulse from a chirped-pulse-amplified mode-locked Ti:sapphire laser system (CPA) is shaped via a Fourier-plane pulse shaper (S) and incident on a cell (C) containing a gas of K_2 molecules at 400°C. Absorption generates a wave packet in the lowest excited electronic state $A^1\Sigma_u^+$ of the molecules. Fluorescence from this state is collected in the near-forward direction by two off-axis parabolic mirrors (P1 and P2) and focused on a nonlinear crystal (X) where it is mixed with a portion of the unshaped laser pulse [the gate pulse, derived from the laser pulse by a beamsplitter (BS)] that has been sent through an adjustable delay line (DL). The resulting frequency-upconverted radiation is imaged (depicted as "A") through a broadband filter (F) onto the slits of a monochromator (MN) after which it is detected by a photomultiplier tube (PMT). Quantum beats are observed in the time- and frequency-resolved fluorescence as the vibrational wave packet oscillates in the excited electronic state potential, as illustrated in the middle panel of (B). The resulting detected signal is shown in the right panel of (B). The time- and frequency-resolved signal shows quantum beats as a function of the time delay τ between the gate pulse and excitation pulse, which have a periodicity $2\pi/\omega$ corresponding to the wave packet oscillation frequency ω. The signal is processed in a computer with a genetic algorithm (GA) and used to reshape the pulse that excites the wave packet in the dimers. This closed-loop control system searches for a wave packet shape that is most immune to the dephasing caused by the rotational-vibrational coupling. The optimal pulse profile (red line) is shown in (B) on the left.

approach is comparable to searching for a DFS for the system, without knowing a priori that such a subspace exists, even in principle. This capability is important for the common situation in which the system-environment coupling is known in insufficient detail to formally address the question.

Among the simplest physical systems that satisfy these requirements is a diatomic molecule (29, 30). Here, the system is the vibrational mode of internuclear motion in the excited electronic state. The wave packet localization is detected by means of time- and frequency-resolved fluorescence that exhibits (for localized states) quantum beats, the visibility of which forms the control signal. The visibility is calculated from the experimental data as the normalized difference of the fluorescence intensity at the peak and at the valley of a specific quantum beat. This CS requires only two data points as input to the control loop, and the value of the visibility ranges between zero (no quantum beats) and unity or 100% (full modulation of the fluorescence intensity). Measurement of the fluorescence projects out only the vibrational motion in the excited electronic state of the molecule. The reservoir is the rotational degree of freedom of the molecule, which is coupled to the vibrational motion through the moment of inertia. Changes in the moment of inertia as the molecule vibrates cause dephasing of the vibrational state. The distribution of population in the reservoir degree of freedom, which at high temperatures forms a quasi-continuum, is unaffected by this coupling to the system. Note that the quantum state of a system, and all state characteristics such as the coherence, indeed describe an ensemble of identical systems. Correspondingly, in the present case, although dephasing of the vibrational wave packet is caused by a purely intramolecular process of the vibration-rotation interaction, the coherence of the vibrational quantum system refers to the ensemble of molecules. This reference to the ensemble is true as well for all quantitative measures of the coherence, which are functions of the quantum state operator, and for the CS, which is deduced from the ensemble average of an observable.

The apparatus used to control the initial shape of the wave packet, and thus to manage the rate at which it dephases, is shown in Fig. 1A. Optical pulses of ~90-fs duration, whose mean wavelength is tuned to the lowest electronic transition of the potassium dimer, are shaped by means of an acousto-optic modulator in a Fourier-plane shaper before exciting a vibrational wave packet in the $^1\Sigma_u^+$ electronic state of the molecules, which were kept in the gas phase at ~400°C. The vibrational wave packet produced by the optical pulse in the excited electronic state has a higher purity than the initial thermal vibrational state in the ground electronic state because of the selectivity of the Franck-Condon transition. The electronically excited molecules vibrate with a period of ~500 fs, and the fluorescence emanating from transitions near the outer turning point of the wave packet's motion exhibits quantum beats with this period (Fig. 1B). The beats decay in amplitude (Fig. 2), as a result of several physical effects. The dominant effect is dephasing resulting from the rotational bath, which has an estimated decay time of 3 ps for this molecule and temperature (30). Delocalization of the wave packet due to the anharmonic character of the vibrational potential also causes a reduction in the beat visibility. This purely Hamiltonian evolution cannot be distinguished from dephasing by the CS, though the purity of the state remains constant for this kind of dynamics. Hence, the CS provides a sufficient but not necessary condition for the existence of coherence: The absence of localization does not necessarily imply that the state of the vibrational mode is completely mixed (corresponding to a vibrational density matrix with zero off-diagonal elements), whereas the presence of localization certainly indicates that that state is not completely mixed. The time scale for anharmonic delocalization is ~12 ps, based on the known molecular potentials (31), which is much longer than the dephasing time due to rotational-vibrational coupling. Collisional dephasing of the vibrations is negligible in a low-density gas at this temperature and pressure. Finally, the spin-orbit coupling of the $^1\Sigma_u^+$ and $^3\Pi_u$ states causes a transfer of population to an electronic state that is not detected by our fluorescence measurements (32). The time scale for this process to occur in the present system is many tens of picoseconds.

The control protocol consists of a measurement of the visibility of the quantum beats near a particular delay. This delay occurs at period 15 of the wave packet oscillations and corresponds to about 3.8 times
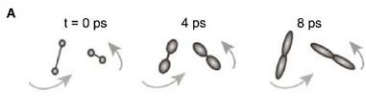
Fig. 2. (A) Illustration of the dephasing of a vibrational wave packet by coupling to rotations. The moment of inertia changes as the localized nuclei move. The resulting modulation of the interaction energy between these degrees of freedom causes the wave packet to dephase and to delocalize. The fluorescence quantum beats then disappear. **(B)** Experimental measurement of the time-resolved fluorescence from K_2 at a wavelength of 1.065 μm, corresponding to the outer turning point of the motion of the vibrational wave packet in the A state. The horizontal axis is the delay between the excitation pulse and the gate pulse used to sample the fluorescence by means of upconversion. The vertical arrow indicates the delay time corresponding to the 15th period of the quantum beats used as the optimization target. The top inset shows the beat pattern for a transform-limited excitation pulse: In this case, the beats decay rapidly as a result of dephasing of the vibrational wave packet, leading to no measurable beat visibility at the target point. The long scan shows the pattern when the optimized laser pulse shape is used, characterized by a slower decay rate of the beats, and a beat amplitude well above the noise at the target point. **(C)** The collected data (circles with the error bars indicating ± SD) show good correspondence with our theoretical model (solid line).
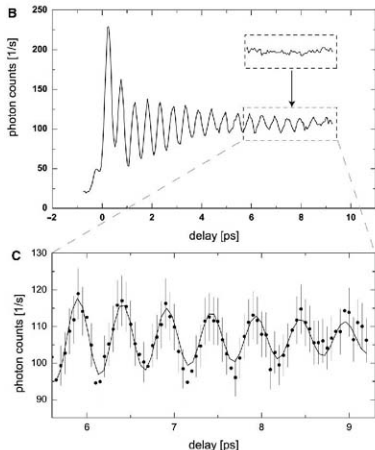
the measured dephasing time τ_d for the vibrational state excited by the transform-limited pulse. Before the start of the search (i.e., for the transform-limited excitation pulse), no quantum beats were visible in the signal at this delay (Fig. 2B, inset). We used an evolutionary algorithm to search through a variety of wave packet shapes (by changing the shape of the excitation pulse) and found an improvement of the quantum-beat visibility from zero to ~7%, or more than four times the noise level. Repeating this experiment for several different time delays (1.9, 3.8, and 5.6 ps) after the excitation pulse produced similar results (with different optimal pulse shapes), showing that the control pulse shape was able to affect the longevity of the localization of the vibrational wave packet. In order to quantify this

effect, and to determine whether controlling the beat visibility was correlated with controlling the decay time of the wave packet's coherence, we measured the decay rate by fitting a damped sinusoid to a full fluorescence beat pattern for each control pulse shape (Fig. 3A). The decay time measured for the optimal control pulse found by the genetic algorithm was about 3.2 ps (i.e., about 1.6 times τ_d). This shows that the details of the pulse shape affect the subsequent dephasing dynamics. Given that the environment is at 400°C, it is perhaps unexpected to find that any major changes in the coherence decay time can be attributed to subtle changes in the initial state of the vibrational mode.

To understand the outcome of the closed-loop search, we characterized the electric field of the

optimal control pulse (33) (Fig. 1B, left). We then performed a series of experiments in which the prominent features of the control pulse fields were examined in turn. For each pulse shape, the decay of the quantum beats was measured and a decay time extracted by fitting the beat pattern to an exponentially damped sinusoid (Fig. 3A). The decay times of the quantum beats for several different pulse shapes are shown in Fig. 3B. First, we examined 10 control pulse shapes with no amplitude modulation but with different values of negative quadratic chirp. The longevity of the quantum beats increased with increasing chirp, the longest decay time [about 4.0 ps (i.e., twice τ_d)] corresponding to the largest negative chirp (-3.1×10^3 fs^2) (Fig. 3B). For pulses with positive chirp, the beat decay time decreased as compared with that of the transform-limited pulse (we examined 10 pulse shapes with different values of positive quadratic chirp). The largest negative chirp leads to a decay time longer than that found via the closed-loop control optimization, although the visibility of the quantum beats around 7.5 ps is smaller. The difference between these outcomes arises because the signal guiding the control loop is the amplitude of the quantum beats at a specified time delay, rather than the decay time extracted from the entire beat pattern. Again, this result is a consequence of our strategy to run the control loop with the minimal amount of data necessary. Measuring the entire decay pattern of quantum beats and fitting the obtained result to a damped sinusoid are good methods for analysis, but this requires a very large number of measurements, across 200 different time delays that take about 30 min to complete per control pulse shape. Therefore, it would be impractical to use such a long sequence of measurements as part of the control loop. On the other hand, the beat visibility at just one time delay yields less information about the system behavior than the entire decay curve, but the visibility measurement can be completed in just 20 s and is therefore well suited for use in a high-bandwidth control loop. We observed that although the strategy to run the control loop using the minimal data necessary did not maximize the decay time, it was still very effective in increasing the longevity of coherence. Next, a control pulse shape consisting of a sequence of subpulses—each shorter than the vibrational period, and each without chirp—was applied to the molecules. A small increase in the decay time was observed when the temporal separation between the subpulses was equal to the vibrational period. We concluded from this analysis that negative chirp was the most important characteristic of the closed-loop search outcome.

The mechanism by which the optimally shaped pulse mitigated the dephasing of the vibrational wave packet was identified by analyzing a model of the system based on a rotational-vibrational coupling Hamiltonian for the excited electronic state of the molecule. Within the Born-Oppenheimer approximation,
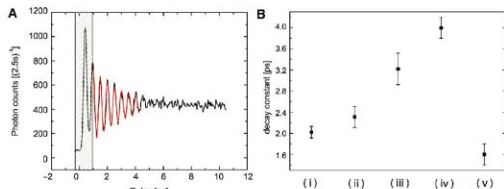
Fig. 3. (**A**) An exponentially damped sine wave (red line) is fitted through the quantum beats outside the control window (cross-hatched box) during which the light field is present. (**B**) The decay time of the quantum beats for several pulse shapes: (i) transform-limited excitation pulses, (ii) pulses found with the genetic algorithm constrained to have flat phase, (iii) pulses found by the genetic algorithm with no constraints, (iv) pulses with -3.1×10^3 fs^2 negative chirp, and (v) pulses with 3.1×10^3 fs^2 positive chirp. The error bars indicate ± SD of the fit.
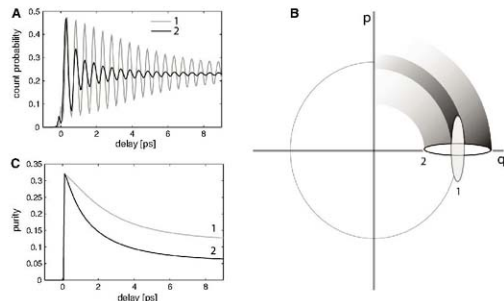
Fig. 4. Illustration of the coherence-preserving mechanism for two states labeled "1" and "2," calculated via the system-bath model. (**A**) The quantum beats decay at different rates for the two states. (**B**) The dephasing spreads the phase-space distribution along the classical trajectory (gray circle). The coherence lifetime is extended if the orientation in phase space of the initial state is along the classical trajectory. (**C**) The total uncertainty in phase space remains smaller for phase-space distribution 1 (representing state 1) than for phase-space distribution 2 (representing state 2), and consequently the purity decays less rapidly.

the Hamiltonian $\hat{H}_i$ of a diatomic molecule in a particular electronic state $|i\rangle$ has the form

$$\hat{H}_i = \frac{\hat{p}^2}{2\mu} + U_i(\hat{q}) + \frac{\hat{J}^2}{2\mu\hat{q}^2} \quad (1)$$

where $\hat{p}$ and $\hat{q}$ are the internuclear momentum and position operators, μ is the reduced mass, $U_i(\hat{q})$ is the adiabatic vibrational potential of the state $|i\rangle$, and $\hat{J}$ is the angular momentum operator of the molecule. This system can be analyzed completely in terms of the known eigenstates or in terms of a system-bath model (30). Both models reveal that in the position-momentum phase space of the vibrational mode, the effect of the rotational dephasing is to smear the vibrational wave packet along the closed classical trajectory of the oscillations. The area of phase space occupied by a quantum state is simply related to its purity. Therefore, vibrational states whose phase-space distributions are predominantly oriented along the classical trajectory will experience a less rapid growth in the uncertainty product $\Delta q \Delta p$, and thus a less rapid decay of purity, than states of similar initial phase space area whose distributions are oriented orthogonal to the classical trajectory (Fig. 4). Experimentally, we find that pulses with negative chirp produce wave packets that are radially squeezed in phase space (34) and thus preserve the purity. Thus, control of the relative phases of the basis states in a coherent quantum superposition provides a natural means to counteract the dephasing process. This simple picture also suggests the limit of possible improvement: If the chirped excitation pulse is as long as one vibrational period (about 500 fs), the initial phase-space distribution is smeared out over 2π, and the purity cannot decay further. However, this completely delocalized state does not show fluorescence quantum beats and so is not controllable via our strategy. The full-model calculation also shows that if the chirp becomes larger, the initial purity is lower than that for the transform-limited pulse, so the optimal solution is a trade-off between a state with very long–lived coherence and one that shows clear localization. Our CS effectively mediates between these limits.

The physical mechanism by which these tailored pulse shapes mitigate dephasing entails control of the fluctuations in the molecules' moment of inertia, which determine the strength of coupling between the rotational and the vibrational degrees of freedom. Negatively chirped pulses generate wave packets that directly reduce the effect of the centrifugal coupling term in Eq. 1. A system-bath model derived from Eq. 1 confirms that the control loop correctly identifies the wave packet shapes that are most resistant to dephasing (29). The model shows that position-squeezed vibrational states are approximate eigenstates of the interaction Hamiltonian. The smaller the position variance immediately after excitation, the less the state is affected by the bath coupling. Although the transform-limited pulse produces the vibrational wave packet with equal position and momentum

uncertainties, negatively chirped optical control pulses generate squeezed-state wave packets whose position variance is decreased at the cost of increased momentum variance.

The second most important component of the closed-loop search outcome was a sequence of subpulses. When consecutive subpulses are separated by one vibrational period, such a sequence creates constructive interference of multiple wave packets at the target time. These interferences increase the dephasing rate only by a small amount and are an epiphenomenon of our choice of CS.

The time-evolution of the purity of vibrational quantum states excited by optical control pulses may be calculated numerically via Eq. 1 by making use of the known rotational-vibrational spectra and the measured electric field of the excitation pulse. This calculation confirms that the fluorescence quantum-beat visibility is a reasonable surrogate for purity, and thus for quantum coherence, because the purity lasts longer for negatively chirped excitation pulses than for the transform-limited pulse. The calculation of the complete ro-vibrational density matrix is numerically expensive and takes typically several hours because of the dimensionality of the combined system-reservoir Hilbert space. Further, maximizing the purity P of the system state is not a convex problem, so it is not possible to guarantee that solutions are globally optimal. As a separate calculation, we replaced the anharmonic vibrational potentials with harmonic ones. The computed quantum-beat visibility and purity show the same qualitative behavior in both cases, and we conclude that the anharmonicity of the adiabatic Born-Oppenheimer potentials does not play an essential role in this experiment. Our measurements do not distinguish between Hamiltonian delocalization and decoherence, and therefore our control strategy would be problematic if anharmonic collapse and rotational dephasing took place on the same time scale; a different CS should be chosen in that case. In other cases where the anharmonicity is large and the anharmonic collapse is rapid as compared with the dephasing, one could optimize the quantum-beat visibility at the first anharmonic full revival. Given the effectiveness of the CS used here, we speculate that any experimentally accessible measure of coherence would form a potential surrogate for use as a control signal. For example, standard time-domain nonlinear optical measurements depend on off-diagonal density matrix elements of the quantum state and thus could also be used as a CS.

We have shown experimentally that control of dephasing by closed-loop methods is a viable approach to reducing the susceptibility of coherent superpositions of quantum states to the effect of an environment. We chose a system sufficiently complex to exhibit generic dephasing dynamics and yet appropriately simple to lend itself to theoretical analysis. Therefore, we were able to numerically simulate the experiment and understand a posteriori the decoherence-preserving mechanism. In more complex systems, this degree of

analysis will not be possible. Also, it will be impossible to conceive of an a priori design strategy, especially if the environmental couplings are not known. In contrast, closed-loop control is almost universally applicable and often the only practical way of controlling the coherence of the quantum system. The key to achieving the coherence preservation is to identify a CS that is simple to measure in the laboratory. A direct determination of the purity of the quantum state requires too much data to make it an effective feedback signal for the control loop. With an easily measurable surrogate, the closed-loop method enables the protection of coherence in a system by successfully identifying states that are most immune to dephasing. We expect that this strategy will also be effective at finding true decoherence-free states, where such states exist. In situations where one or more decoherence-free states are available, the search algorithm will be able to locate them, again without the use of previous knowledge of the system-environment couplings. The outcome of our experiment is that appropriately constructed wave packets containing more coherence—in the sense of a larger number of off-diagonal elements of the density matrix in the bare system basis—preserve their coherence longer in the presence of environmental noise than wave packets with fewer populated basis states. Of course, a correct phasing of the states in the quantum superposition is crucial for rendering the wave packet less susceptible to dephasing. Thus, it appears that coherence may be used to fight decoherence: a rather unexpected result, but one that may have implications for other quantum systems and for applications demanding control of quantum interferences.

References and Notes
1. R. S. Judson, H. Rabitz, *Phys. Rev. Lett.* **68**, 1500 (1992).
2. T. C. Weinacht, J. Ahn, P. H. Bucksbaum, *Nature* **397**, 233 (1999).
3. J. L. Herek, W. Wohlleben, R. J. Cogdell, D. Zeidler, M. Motzkus, *Nature* **417**, 533 (2002).
4. T. Brixner *et al.*, *Phys. Rev. Lett.* **92**, 208301 (2004).
5. A. Lindinger *et al.*, *Phys. Rev. Lett.* **93**, 033001 (2004).
6. V. I. Prokhorenko *et al.*, *Science* **313**, 1257 (2006).
7. R. Bartels *et al.*, *Nature* **406**, 164 (2000).
8. R. J. Levis, G. M. Menkir, H. Rabitz, *Science* **292**, 709 (2001).
9. T. Laarmann *et al.*, *Phys. Rev. Lett.* **98**, 058302 (2007).
10. H. Lee, Y. Cheng, G. R. Fleming, *Science* **316**, 1462 (2007).
11. A. Ekert, R. Jozsa, *Rev. Mod. Phys.* **68**, 733 (1996).
12. A. Ekert, R. Jozsa, *Philos. Trans. R. Soc. London Ser. A* **356**, 1769 (1998).
13. A. M. Steane, *Phys. Rev. Lett.* **77**, 793 (1996).
14. J. Preskill, *Proc. R. Soc. London Ser. A* **454**, 385 (1998).
15. D. Vitali, P. Tombesi, *Phys. Rev. A* **59**, 4178 (1999).
16. E. A. Shapiro, I. A. Walmsley, M. Yu. Ivanov, *Phys. Rev. Lett.* **98**, 050501 (2007).
17. D. A. Lidar, L. L. Chuang, K. B. Whaley, *Phys. Rev. Lett.* **81**, 2594 (1998).
18. D. J. Tannor, A. Bartana, *J. Phys. Chem. A* **103**, 10359 (1999).
19. S. E. Sklarz, D. J. Tannor, N. Khaneja, *Phys. Rev. A* **69**, 053408 (2004).
20. U. Hohenester, G. Stadler, *Phys. Rev. Lett.* **92**, 196801 (2004).
21. U. Hohenester, *Phys. Rev. B* **74**, 161307(R) (2006).
22. H. Jirari, W. Pötz, *Phys. Rev. A* **74**, 022306 (2006).
23. G. Katz, M. A. Ratner, R. Kosloff, *Phys. Rev. Lett.* **98**, 203006 (2007).

24. M. Grace et al., J. Phys. B 40, S103 (2007).
25. T. Hansson, Phys. Rev. A 61, 033404 (2000).
26. F. A. Bovino et al., Phys. Rev. Lett. 95, 240407 (2005).
27. T. J. Dunn, I. A. Walmsley, S. Mukamel, Phys. Rev. Lett. 74, 884 (1995).
28. T. J. Dunn, J. N. Sweetser, I. A. Walmsley, C. Radzewicz, Phys. Rev. Lett. 70, 3388 (1993).
29. C. Brif, H. Rabitz, S. Wallentowitz, I. A. Walmsley, Phys. Rev. A 63, 063404 (2001).
30. S. Wallentowitz, I. A. Walmsley, L. J. Waxer, Th. Richter, J. Phys. B 35, 1967 (2002).

31. A. J. Ross, P. Crozet, C. Effantin, J. D'Incan, R. F. Barrow, J. Phys. B 20, 6225 (1987).
32. I. A. Walmsley, L. Waxer, J. Phys. B 31, 1825 (1998).
33. C. Iaconis, I. A. Walmsley, Opt. Lett. 23, 792 (1998).
34. J. Janszky, A. V. Vinogradov, I. A. Walmsley, J. Mostowski, Phys. Rev. A 50, 732 (1994).
35. This work was supported by the Defense Advanced Research Projects Agency (DARPA) QuIST. I.A.W. acknowledges support by the UK Quantum Information Processing Interdisciplinary Research Collaboration (funded by the Engineering and Physical Sciences

Strong Interactions in Multimode Random Lasers

Hakan E. Türeci,[1]* Li Ge,[2] Stefan Rotter,[2]† A. Douglas Stone[2]

Unlike conventional lasers, diffusive random lasers (DRLs) have no resonator to trap light and no high-Q resonances to support lasing. Because of this lack of sharp resonances, the DRL has presented a challenge to conventional laser theory. We present a theory able to treat the DRL rigorously and provide results on the lasing spectra, internal fields, and output intensities of DRLs. Typically DRLs are highly multimode lasers, emitting light at a number of wavelengths. We show that the modal interactions through the gain medium in such lasers are extremely strong and lead to a uniformly spaced frequency spectrum, in agreement with recent experimental observations.

Novel laser systems have emerged recently because of modern nanofabrication capabilities (1–3). The diffusive random laser (DRL), perhaps the most challenging of the new systems, consists of a random aggregate of particles that scatter light and have gain or are embedded in a background medium with gain (2, 4–8). Whereas light scattering in such a random medium can give rise to Anderson-localized, high-Q resonances (9, 10), in almost all experiments the localized regime is not reached, and the laser "cavity" has no isolated resonances in the absence of gain. Despite the lack of sharp resonances, the laser emission from the more recent DRLs (2, 5, 6) was observed to have the essential properties of conventional lasers: the appearance of coherent emission with line-narrowing above a series of thresholds and uncorrelated photon statistics far from threshold (11). Earlier work on random lasers (4, 7) did not find isolated narrow lines and was interpreted as incoherent lasing, in which there was intensity feedback but not amplitude feedback. Later experiments (2) and recent numerical studies (12) indicated that the lasing involves coherent phase-sensitive feedback in at least some cases. Our work shows that standard coherent multimode lasing is possible even when the linear resonances are much broader than their spacing, raising the question of what determines the emission frequencies of DRLs because they are not determined by the

position of passive cavity resonances. Furthermore, recent experiments on porous GaP DRLs have shown that the frequencies are rather uniformly spaced and stable from pulse to pulse, although the intensities vary substantially (8). We show that this is a result of strong nonlinear interactions between lasing modes combined with extreme leakiness, a regime particularly difficult to treat. In any multimode laser, the different modes compete with one another through the gain medium in a complex manner that depends on the spatial distribution of the electric field of each mode. This is particularly severe in the DRL, in which there are many spatially overlapping modes with similar (very short) lifetimes.

The finesse, f, of a resonator is defined as the ratio of the resonance spacing to the resonance width; standard laser theory only addresses lasers with high-finesse (weakly open) resonators and cannot be applied to the DRL, which has finesse much less than unity. Hence, no analytic results have been derived relating to two-dimensional (2D) or 3D DRLs, and realistic numerical simulations have been limited because of the computational demands. We introduced a formulation of semiclassical laser theory in terms of biorthogonal modes, called constant flux (CF) states, which treats lasing media with any degree of outcoupling and includes the effects of nonlinear modal interactions to all orders (13, 14). We present analytic and numerical results using this approach applied to a DRL.

The simplest model for a laser that captures all of the relevant spatial complexity uses the Maxwell-Bloch equations (15, 16), which are three coupled nonlinear equations for the electric field, the polarization, and the inversion of the gain medium. For stationary multimode lasing, the modes predicted by these equations are deter-

mined by the time-independent self-consistent equation (13)

$$\Psi_\mu(x) = \frac{\dot{r}\gamma_\perp}{\gamma_\perp - i(k_\mu - k_a)}$$

$$\int dx' \frac{D_0(x') G(x,x';k_\mu) \Psi_\mu(x')}{\varepsilon(x')(1 + \sum_\nu \Gamma_\nu |\Psi_\nu(x')|^2)} \quad (1)$$

where the electric field is given by $e(x,t) = \sum_\mu \Psi_\mu(x) e^{-i\Omega_\mu t}$. In Eq. 1, the lasing frequencies $\Omega_\mu = ck_\mu$, and the lasing mode functions $\Psi_\mu(x)$ are assumed to be unknown (henceforth we set the speed of light $c = 1$) and use the wave vector to denote frequency as well). In Eq. 1, k_a is the atomic frequency, $\gamma_\perp$ is the transverse relaxation rate, $\Gamma_\nu = \Gamma(k_\nu)$ is the gain profile evaluated at k_ν, $D_0(x) = D_0 [1 + d_0(x)]$ is the pump, which can vary in space, and $\varepsilon(x) = n^2(x)$ is the dielectric function of the "resonator." Electric field and pump strength are dimensionless, being measured in units $e_c = \hbar \sqrt{\gamma_\perp \gamma_\parallel}/2g$ and $D_{0c} = 4\pi k_a^2 g^2/\hbar\gamma_\perp$, where $\gamma_\parallel$ is the longitudinal relaxation rate and g is the dipole matrix element of the gain medium. Each lasing mode Ψ_μ depends nonlinearly on all of the other lasing modes through the denominator in Eq. 1, which represents the "spatial hole-burning" (15) interaction with the other modes. Through this mechanism, modes that lase first tend to suppress lasing in other modes, particularly those with which they are correlated in space.

For simplicity we study a 2D DRL and take the gain medium to be a uniform disk of radius R, which contains randomly placed nanoparticles with constant index greater than unity. The light field in the cavity can be either transverse magnetic or transverse electric polarized perpendicular to the plane of the disk, leading to a scalar equation for its normal component. The integral in Eq. 1 is over the gain region, and the kernel $G(x,x';k)$ is the Green function of the cavity wave equation with purely outgoing boundary conditions (13). This represents the steady-state response of the passive cavity to a monochromatic dipole oscillating with frequency k at x'. The nonhermitian boundary conditions on the Green function lead to a spectral representation $G(x,x';k)$ in terms of dual sets of biorthogonal functions $\varphi_m(x, k)$ and $\overline{\varphi}_m(x, k)$, termed constant flux (CF) states, with complex eigenvalues, k_m (13).

The CF states play the role of the linear cavity resonances within our theory and reduce to the quasi-bound (QB) states within the cavity to a

[1]Institute of Quantum Electronics, Eidgenössische Technische Hochschule (ETH) Zürich, 8093 Zürich, Switzerland. [2]Department of Applied Physics, Post Office Box 208284, Yale University, New Haven, CT 06520-8284, USA.

*To whom correspondence should be addressed. E-mail: tureci@phys.ethz.ch
†Presently on leave from Technische Universität Wien, Wiedner Hauptstraße 8-10/136, A-1040 Vienna, Austria.
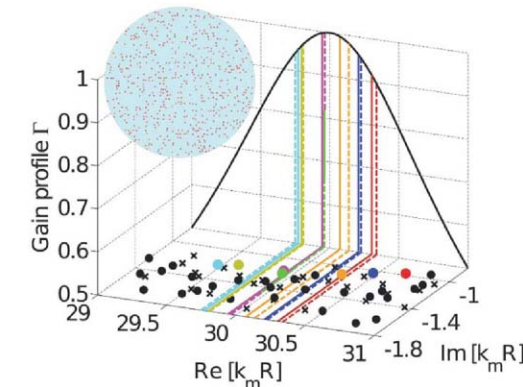
Fig. 1. Lasing frequencies of a DRL. Black circles and crosses represent the complex frequencies of the CF and QB states, respectively; they are distinct but statistically similar. Because their spacing on the real axis is much less than their distance from the axis, the system has no isolated linear resonances, and the cavity has average finesse less than unity ($f \approx 0.05$). Solid colored lines represent the actual frequencies k_μ of the lasing modes at pump $D_0/D_{0c} = 123.5035$; dashed lines denote the values $k_\mu^{(0)}$ arising from the single largest CF state contributing to each mode (the CF frequency is denoted by the corresponding colored circle). The thick black line represents the atomic gain curve $\Gamma(k)$, peaked at the atomic frequency, $k_a R = 30$. Because the cavity is very leaky, the lasing frequencies are strongly pulled to the line center in general; however, the collective contribution to the frequency is random in sign. (**Inset**) Schematic of the configuration of nanoparticle scatterers in the disk-shaped gain region of the DRL.

Fig. 2. Lasing intensities of a DRL. Modal intensities of a DRL versus pump D_0 for the disorder configuration of Fig. 1, illustrating complex nonlinear dependence and modal interactions. For example, the black and orange modes approach each other in frequency and interact so strongly that the black mode is driven to zero, at which point the orange mode has a kink in its intensity curve. (**Inset**) The frequencies of each of the eight lasing modes both above (solid) and below (dashed) threshold ver-

sus D_0. Note that frequencies can cross if one mode is not lasing but not if both are lasing, and modes with nearby frequencies interact strongly so that their intensity curves are highly anticorrelated. The interactions increase the lasing thresholds substantially. For example, the green and purple modes with $k_\mu = 29.95$ have noninteracting thresholds 73.4460 and 75.6919, respectively, but the hole-burning interaction pushes that of the purple mode up to $D_0/D_{0c} = 105$. In addition there would be 16 modes lasing by $D_0/D_{0c} = 93$ in the absence of interactions, compared to the 6 we find.
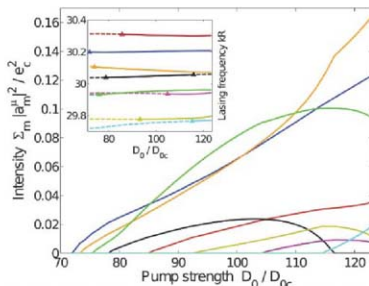
good approximation for a high-Q resonator (17). Importantly, the CF states are complete within the cavity and generate a conserved photon flux outside the cavity, unlike the QB states (13); for the extremely leaky cavity of a DRL, the CF and QB states are significantly different but statistically similar (Fig. 1).

Because the CF basis is complete and conserves flux outside the gain region, it is an appropriate basis for representing arbitrary lasing modes $\Psi_\mu(x)$ of a DRL. To solve Eq. 1, we expand each mode in terms of CF states: $\Psi_\mu(x) = \sum_{m=1}^{N_{CF}} a_m^\mu \varphi_m(x)$. Substituting this expansion into Eq. 1 gives an equation for the complex vector of coefficients $\mathbf{a}^\mu = (a_1^\mu, a_2^\mu, ..., a_{N_{CF}}^\mu)$ that completely determines Ψ_μ:

$$a_m^\mu = D_0 \sum_n T_{mn}^\mu a_n^\mu \qquad (2)$$

The nonlinear operator T_{mn}^μ is written out explicitly and discussed in (18).

This formalism allows us to obtain analytic insight into the question of what determines the frequencies of the DRL. In single-mode lasers, each lasing frequency is a weighted mean of the real part of the cavity resonance frequency and the atomic frequency (16), which for a typical high-finesse system is very close to the cavity frequency with a small shift ("pull") toward the atomic line. If we denote the "conventional" lasing frequency by $k_\mu^{(0)}$, we find from Eq. 2 that for the DRL

$$k_a \approx k_\mu^{(0)} + k_\mu^{(c)} \qquad (3)$$

where $k_\mu^{(c)}$ is a collective contribution due to all the other CF states, which has no analog in conventional lasers. In our parameter regime ($k_a R \approx 30$), both the conventional and collective terms are important (although the conventional term is larger), and the lasing frequencies have no simple relationship to the cavity frequencies. The collective term is random in sign and does not always generate a pull toward line center (Fig. 1). We believe that at larger $k_a R$ the collective term will dominate.

In Fig. 2, we plot the intensities associated with the lasing modes of Fig. 1 as a function of pump strength, D_0, measured by the length of the vector of CF coefficients, $I = \sum_{m=1}^{N_{CF}} |a_m^\mu|^2$. The behavior is very different from conventional lasers, showing complex nonmonotonic and reentrant behavior in contrast to the linear increase found for uniform edge-emitting lasers (14). Analysis reveals that the complex behavior is due to the strong spatial hole-burning interactions in these systems. The Fig. 2 inset shows the lasing frequencies associated with the modes as a function of pump; the eight lasing modes in the interval, there are six that form three pairs nearby in frequency, and their behavior is highly correlated. Evaluation of the overlap of the $\mathbf{a}^\mu$ vector associated with each pair of modes confirms that not only their frequencies but also their decomposition into CF states are similar.

Equations 1 and 2 imply that modes with similar $\mathbf{a}^\mu$ vectors and similar frequencies will
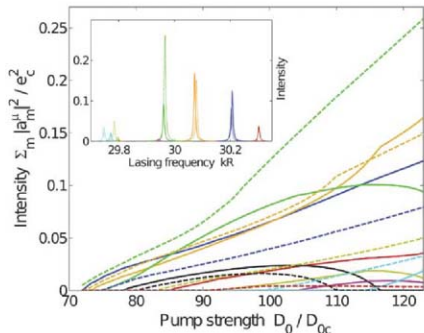
Fig. 3. Intensity and frequency fluctuations in a DRL. Comparison of modal intensities of the DRL for the same disorder configuration in the case of uniform pumping (solid lines) and partially nonuniform pumping (dashed lines). The main source of the large intensity fluctuations is the shift in thresholds. This has the largest effect for nearly degenerate mode pairs such as the green and purple modes ($k_L \approx 29.95$). (**Inset**) Lasing spectra at $D_0/D_{0c} = 123.5035$ (lines broadened for visibility). Note that the black mode does not appear at this pump because it has already been suppressed by the orange mode, and the purple mode only appears for uniform pumping because it never reaches threshold in the nonuniform case. The intensities at this pump can fluctuate by more than a factor of two between the two cases, whereas the frequencies fluctuate by just a few percent of their average spacing.
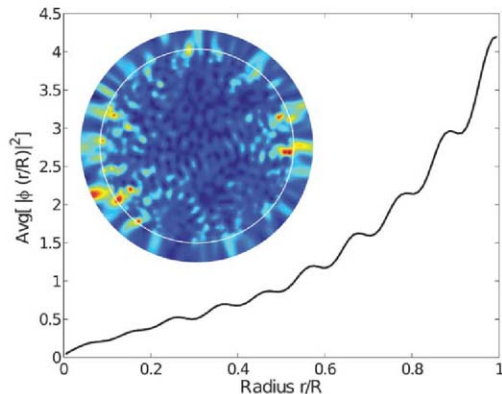
Fig. 4. Field distribution of a DRL. Radial intensity of CF states contributing to the lasing modes averaged over 400 disorder configurations. There is a large nonrandom increase of intensity with radius r. (**Inset**) False color plot of electric field intensity of the seven lasing modes of the DRL of Fig. 1 at pump $D_0/D_{0c} = 123.5035$ (white circle is boundary of gain medium). Note brightest regions appear at the edge of the gain medium; this is characteristic of low finesse lasers but is a particularly large effect in the DRL.

compete strongly because this leads to a hole-burning denominator that is spatially correlated with the numerator. However, it is not obvious that frequency quasi-degeneracy should be associated with spatial correlation for the DRL. For random lasers with Anderson localization (9, 10), the CF states would typically be spatially disjoint, the $T_{mn}(k)$ operator (compare with Eq. S3) would be approximately diagonal, and there would be no such spatial correlation. In additional calculations not shown, we do find that for larger index nanoparticles, which begin to localize the CF states, the modal interactions are reduced. But for the DRL, $T_{mn}(k)$ is not diagonal, and frequency degeneracy would require an eigenvalue degeneracy in this complex pseudo-random matrix [see discussion in (18)]. Instead, there is eigenvalue repulsion in the complex plane and strong mixing of eigenvectors, resulting in large spatial overlap of quasi-degenerate lasing modes and strong hole-burning interaction. This interaction, in the absence of some special symmetry, tends to suppress one of the two modes, leading to well-spaced lasing frequencies as found by (8). Hence, the rigid lasing-frequency spectrum could distinguish the DRL from an Anderson-localized laser.

This strong interaction of nearly degenerate modes is reflected in the lasing threshold of the second mode of each pair (Fig. 2 caption) (18). These interaction effects are strongly nonlinear and hence highly sensitive to statistical fluctuations. To illustrate this, in Fig. 3 we contrast the intensity behavior of Fig. 2, for which the pump was uniform in space [$d_0(x) = 0$], with a case for which we have added to the uniform pump a random white noise term $d_0(x)$ of standard deviation ±30% (normalized to the same total power). For this nonuniform pump, the third uniform mode (green) now turns on first. It is thus able to suppress the seventh uniform mode (purple) over the entire range of pump powers and acquires an intensity almost a factor of 3 greater at the same average pump power. The intensities of all the interacting pairs show similar high sensitivity to pump profile, whereas their frequencies remain relatively stable (Fig. 3 inset). Exactly such behavior was observed in shot-to-shot spectra of DRLs in experiments (8).

Lastly, we consider the spatial variation of the electric field in DRLs (Fig. 4). The false-color representation of the multimode electric field in the laser has a striking property: It is consistently brighter at the edge of the disk than at its center, even though the gain is uniform and there are no special high-Q modes localized near the edge. To demonstrate that this effect is not a statistical fluctuation associated with this particular disorder configuration, we have averaged the behavior of the entire basis set of CF states over disorder configurations. The result is a nonrandom average growth of intensity toward the boundary. The origin of this effect is known from earlier work on distributed feedback lasers with weak reflectivity (19); if the single-pass loss is large,

then the single-pass gain must also be large in order to lase, leading to a visible nonuniformity of the lasing mode, with growth in the direction of the loss boundary (on average the radial direction for the DRL). Because the DRL has fractional finesse (which is not achievable in a 1D geometry), this effect is much larger in these systems and should be observable. This effect means that the electric field fluctuations in DRLs will differ substantially from the random matrix/quantum chaos fluctuations of linear cavity modes (20), first because each mode is a superposition of pseudo-random CF states and second because these CF states themselves are not uniform on average.

The coexistence of gain, nonlinear interactions, and overlapping resonances (fractional finesse) makes the DRL a more complex and richer system than the widely studied linear wavechaotic systems. It remains to be seen whether concepts from random matrix theory and semiclassical quantum mechanics (quantum chaos) will prove fruitful in this context. The theory

presented here is ab initio in the sense that it generates all properties of the lasing states from knowledge of the dielectric function of the host medium and basic parameters of the gain medium; it should be applicable to any novel laser-cavity system.

References and Notes
1. K. J. Vahala, *Nature* **424**, 839 (2003).
2. H. Cao, *J. Phys. A* **38**, 10497 (2005).
3. O. Painter *et al.*, *Science* **284**, 1819 (1999).
4. N. M. Lawandy, R. M. Balachandran, A. S. L. Gomes, E. Sauvain, *Nature* **368**, 436 (1994).
5. S. Mujumdar, M. Ricci, R. Torre, D. S. Wiersma, *Phys. Rev. Lett.* **93**, 053903 (2004).
6. H. Cao *et al.*, *Phys. Rev. Lett.* **82**, 2278 (1999).
7. H. Cao, *Waves Random Media* **13**, R1 (2003).
8. K. L. van der Molen, R. W. Tjerkstra, A. P. Mosk, A. Lagendijk, *Phys. Rev. Lett.* **98**, 143901 (2007).
9. P. Pradhan, N. Kumar, *Phys. Rev. B* **50**, 9644 (1994).
10. V. Milner, A. Z. Genack, *Phys. Rev. Lett.* **94**, 073901 (2005).
11. H. Cao, Y. Ling, J. Y. Xu, C. Q. Cao, P. Kumar, *Phys. Rev. Lett.* **86**, 4524 (2001).
12. C. Vanneste, P. Sebbah, H. Cao, *Phys. Rev. Lett.* **98**, 143902 (2007).
13. H. E. Türeci, A. D. Stone, B. Collier, *Phys. Rev. A* **74**, 043822 (2006).
14. H. E. Türeci, A. D. Stone, L. Ge, *Phys. Rev. A* **76**, 013813 (2007).
15. H. Haken, H. Sauermann, *Z. Phys.* **173**, 261 (1963).
16. H. Haken, *Light: Laser Dynamics* (North-Holland, Amsterdam, 1985), vol. 2.
17. In a high-finesse cavity, QB and CF states are essentially the same, each lasing mode is associated with one QB state, and the lasing frequency takes the value $k_\mu^{(0)}$ between the atomic line center and the cavity frequency q_λ determined by the ratio of $\gamma_\perp/\kappa$ (13).
18. Materials and methods are available on *Science* Online.
19. H. Kogelnik, C. V. Shank, *J. Appl. Phys.* **43**, 2327 (1972).
20. A. Kudrolli, S. Sridhar, A. Pandey, R. Ramaswamy, *Phys. Rev. E* **49**, R11 (1994).
21. This work was supported by NSF grant DMR-0408636, by the Max Kade and W. M. Keck foundations, and by the Aspen Center for Physics. We thank R. Tandy, M. Madadha, H. Cao, A. Lagendijk, P. Sebbah, C. Vanneste, and D. Wiersma for discussions.

Supporting Online Material
www.sciencemag.org/cgi/content/full/320/5876/643/DC1
Materials and Methods
Figs. S1 to S3
References and Notes

16 January 2008; accepted 25 March 2008
10.1126/science.1155311

Silica-on-Silicon Waveguide Quantum Circuits

Alberto Politi, Martin J. Cryan, John G. Rarity, Siyuan Yu, Jeremy L. O'Brien*

Quantum technologies based on photons will likely require an integrated optics architecture for improved performance, miniaturization, and scalability. We demonstrate high-fidelity silica-on-silicon integrated optical realizations of key quantum photonic circuits, including two-photon quantum interference with a visibility of 94.8 ± 0.5%; a controlled-NOT gate with an average logical basis fidelity of 94.3 ± 0.2%; and a path-entangled state of two photons with fidelity of >92%. These results show that it is possible to directly "write" sophisticated photonic quantum circuits onto a silicon chip, which will be of benefit to future quantum technologies based on photons, including information processing, communication, metrology, and lithography, as well as the fundamental science of quantum optics.

Quantum information science (1) has shown that quantum mechanical effects can dramatically improve performance for certain tasks in communication, computation, and measurement. Of the various physical systems being pursued, single particles of light (photons) have been widely used in quantum communication (2), quantum metrology (3–5), and quantum lithography (6) settings. Low noise (or decoherence) also makes photons attractive quantum bits (or qubits), and they have emerged as a leading approach to quantum information processing (7).

In addition to single-photon sources (8) and detectors (9), photonic quantum technologies require sophisticated optical circuits involving high-visibility classical and quantum interference.

Centre for Quantum Photonics, H. H. Wills Physics Laboratory and Department of Electrical and Electronic Engineering, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK.

*To whom correspondence should be addressed. E-mail: Jeremy.OBrien@bristol.ac.uk

Although a number of photonic quantum circuits have been realized for quantum metrology (3, 4, 10–13), lithography (6), quantum logic gates (14–20), and other entangling circuits (21–23), these demonstrations have relied on large-scale (bulk) optical elements bolted to large optical tables, thereby making them inherently unscalable.

We demonstrate photonic quantum circuits using silica waveguides on a silicon chip. The monolithic nature of these devices means that the correct phase can be stably realized in what would otherwise be an unstable interferometer, greatly simplifying the task of implementing sophisticated photonic quantum circuits. We fabricated hundreds of devices on a single wafer and find that performance across the devices is robust, repeatable, and well understood.

A typical photonic quantum circuit takes several optical paths or modes (some with photons, some without) and mixes them together in a linear optical network, which in general consists of nested classical and quantum interferometers (e.g., Fig. 1C). In a standard optical implementation, the photons propagate in air, and the circuit is constructed from mirrors and beam splitters (BSs), or half-reflective mirrors, which split and recombine optical modes, giving rise to both classical and quantum interference. High-visibility quantum interference (24) demands excellent optical mode overlap at a BS, which requires exact alignment of the modes, whereas high visibility classical interference also requires subwavelength stability of optical path lengths, which often necessitates the design and implementation of sophisticated stable interferometers. Combined with photon loss, interference visibility is the major contributor to optical quantum circuit performance.
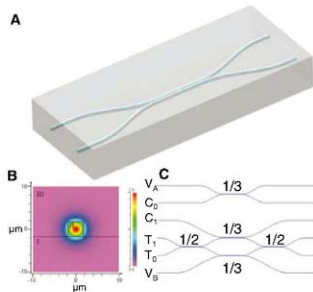
In conventional (or classical) integrated optics devices, light is guided in waveguides—consisting of a core and slightly lower-refractive index cladding (analogous to an optical fiber)—which are usually fabricated on a semiconductor chip. By careful choice of core and cladding dimensions and refractive index difference, it is possible to design such waveguides to support only a single transverse mode for a given wavelength range. Coupling between waveguides, to realize BS-like operation, can be achieved when two waveguides are brought sufficiently close together that the evanescent fields overlap; this is known as a directional coupler. By lithographically tuning the separation between the waveguides and the length of the coupler, the amount of light coupling from one waveguide into the other (the coupling ratio $1 - \eta$, where η is equivalent to BS reflectivity) can be tuned.

The most promising approach to photonic quantum circuits for practical technologies appears to be realizing integrated optics devices that operate at the single-photon level. Key require-

Fig. 1. Silica-on-silicon integrated quantum photonic circuits. (**A**) A directional coupler, which can be used as the building block for integrated photonic quantum circuits by replacing the bulk BS. (**B**) The modeled transverse intensity profile of the guided mode superimposed on the waveguide structure. (**C**) Design of the integrated two-photon CNOT quantum logic gate.

Fig. 2. Quantum interference in an integrated waveguide coupler. The plot shows the rate of detecting a photon at each output of the coupler as a function of the relative delay in arrival time of the photons. The error bars are smaller than the data points.

Fig. 3. Two-photon quantum interference on-chip. (**A**) Quantum interference visibility at 1/2 and 1/3 couplers that compose a CNOT gate (where the 1/2 couplers range from η (1/2) = 0.4 to 0.6 and the 1/3 couplers are 2/3 this value: η (1/3) = 0.27 to 0.4). The fit to the 1/2 data includes an offset in the coupling ratio $\delta\eta$ and mode mismatch ε as free parameters. The same values are used for the 1/3 theoretical curve. (**B**) The average of the logical basis fidelities F for each of the CNOT gates. The solid curve corresponds to a model including only the above values of ε and $\delta\eta$. The model does not include the effect of classical interference, which explains the offset.

ments are single-mode guiding of single photons, high-visibility classical interference, high-visibility quantum interference, and the ability to combine these effects in a waveguide optical network.

We required a material system that (i) is low loss at a wavelength of $\lambda \sim 800$ nm, where commercial silicon avalanche photodiode single-photon counting modules (SPCMs) are near their peak efficiency of ~70%; (ii) enables a refractive index contrast $\Delta = (n_{core}^{2} - n_{cladding}^{2})/2n_{core}^{2}$ that results in single-mode operation for waveguide dimensions comparable to the core size of conventional single-mode optical fibers at ~ 800 nm (4 to 5 μm), to allow good coupling of photons to fiber-coupled single-photon sources and detectors; and (iii) is amenable to standard optical lithography fabrication techniques. The most promising material system to meet these requirements was silica (silicon dioxide SiO_2), with a low level of doping to control the refractive index, grown on a silicon substrate (Fig. 1B).

A refractive index contrast of $\Delta = 0.5\%$ was chosen to give single-mode operation at 804 nm for 3.5 by 3.5 μm waveguides (25). This value of Δ provides moderate mode confinement (the transverse intensity profile is shown in Fig. 1B), thereby minimizing the effects of fabrication or modeling imperfections. We designed a number of devices, including directional couplers with various η's, Mach-Zender interferometers (consisting of two directional couplers), and more sophisticated devices built up from several directional couplers with different η's.
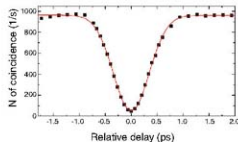
Starting with a 4″ silicon wafer, a 16-μm layer of thermally grown undoped silica was deposited as a buffer (material I in Fig. 1B), followed by flame hydrolysis deposition of a 3.5-μm waveguide core of silica doped with germanium and boron oxides (II). The core material was patterned into 3.5-μm-wide waveguides with standard optical lithography techniques and finally overgrown with a further 16-μm cladding layer of phosphorus and boron-doped silica with a refractive index matched to that of the buffer

(III). The wafer was diced into several dozen individual chips, each containing typically several devices. Some chips were polished to enhance coupling in and out of the waveguides (26).
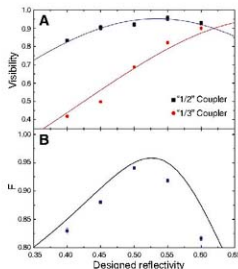
We used a beta-barium borate type-I spontaneous parametric down-conversion (SPDC) crystal, pumped with a 60-mW, 402-nm continuous wave diode laser to produce 804-nm degenerate photon pairs at a detected rate of 4000 s^{-1} when collected into single-mode polarization maintaining fibers (PMFs). We used 2-nm interference filters to ensure good spectral indistinguishability (27). Single photons were launched into the waveguides on the integrated optical chips and then collected at the outputs using two arrays of 8 PMFs, with 250 μm spacing, to match that of the waveguides, and detected with fiber-coupled SPCMs. The PMF arrays and chip were directly butt-coupled, with index matching fluid. Overall coupling efficiencies of ~60% through the device (insertion loss = 40%) were routinely achieved (28).

Figure 2 shows the classic signature of quantum interference: a dip in the rate of detecting two photons at each output of a directional coupler near zero delay in relative photon arrival time (24). The raw visibility (29) $V = 94.8 \pm 0.5\%$ is a measure of the quality of the interference and demonstrates very good quantum behavior of photons in an integrated optics architecture.

Figure 3A shows the measured nonclassical visibility for 10 couplers on a single chip with a range of design η's. The observed behavior is well explained by the theoretical curves, which include a small amount of mode mismatch ε and an offset of $\delta\eta = 3.4 \pm 0.7\%$ from the design ratio. It is inherently difficult to identify in which degree of freedom this small mode mismatch occurs (30). Misalignment of PMF fibers in the array (specified to be <3°) would cause polarization mode mismatch. Small spatial mode mismatch could arise if weakly guided higher-order modes propagate across the relatively short devices (31). These results demonstrate the

high yield and excellent reproducibility of the devices.

General photonic quantum circuits require both quantum and classical interference and their combination for conditional phase shifts (32). An ideal device for testing all of these requirements is the entangling controlled-NOT (CNOT) logic gate shown in Fig. 1C (33, 34), which has previously been experimentally demonstrated using bulk optics (15–19). The control C and target T qubits are each encoded by a photon in two waveguides, and the success of the gate is heralded by detection of a photon in both the control and target outputs, which happens with proba-

bility 1/9. The waveguide implementation of this gate is essentially a direct writing onto the chip of the theoretical scheme presented in (33); it is composed of two 1/2 couplers and three 1/3 couplers.

To allow for possible design and fabrication imperfections, we designed and fabricated on the same chip several CNOT devices with 1/2 couplers ranging from η (1/2) = 0.4 to 0.6 and, correspondingly, 1/3 couplers ranging from η (1/3) = 0.27 to 0.4 (i.e., 2/3 of the 1/2 couplers). The quantum interference measurements described above (Fig. 3B) show that the devices are in fact very close to the design η: $\delta\eta = 3.4 \pm 0.7\%$. To measure the 1/2 couplers, we sent single photons into the T_0 and T_1 inputs and collected photons from the C_1 and V_B outputs (and the reverse for the other 1/2 coupler); the 1/3 data are for the couplers between the C_0 and V_A waveguides (see Fig. 1C).

For the CNOT device with nominally η (1/2) = 0.5 and η (1/3) = 0.33 couplers, we input the four computational basis states $|0\rangle_C|0\rangle_T$, $|0\rangle_C|1\rangle_T$, $|1\rangle_C|0\rangle_T$ and $|1\rangle_C|1\rangle_T$ and measured the probability of detecting each of the computational basis states at the output (Fig. 4A). The excellent agreement for the $|0\rangle_C$ inputs (peak values of 98.5%) is a measure of the classical interference in the target interferometer and demonstrates that the waveguides are stable on

a subwavelength scale—a key advantage arising from the monolithic nature of an integrated optics architecture. The average of the logical basis fidelities (14–20) is $F = 94.3 \pm 0.2\%$. The fidelities for the other four devices (with different η's) are lower (Fig. 3B), as expected.

To directly confirm coherent quantum operation and entanglement in our devices, we launched pairs of photons into the T_0 and T_1 waveguides. This state should ideally be transformed at the first 50:50 coupler as follows:

$$|11\rangle_{T_0 T_1} \rightarrow (|20\rangle_{T_0 T_1} - |02\rangle_{T_0 T_1})/\sqrt{2} \quad (1)$$

that is, a maximally path-entangled superposition of two photons in the top waveguide and two photons in the bottom waveguide. A very low rate of detecting a pair of photons at the C_1 and V_A outputs, combined with a high rate of detecting two photons in either of these outputs (with a pair of cascaded SPCMs) confirmed that the state was predominantly composed of $|20\rangle$ and $|02\rangle$ components but did not indicate a coherent superposition. At the second 50:50 coupler between the T_0 and T_1 waveguides, the reverse transformation of Eq. 1 should occur, provided the minus superposition exists. A high rate of detecting photon pairs at the T_0 and T_1 outputs, combined with a low rate of detecting

two photons in either of these outputs, confirmed this transformation. From each of these measured count rates, we were able to estimate the two-photon density matrix (Fig. 4D). The fidelity with the maximally path-entangled state $|20\rangle - |02\rangle$ is >92% (35). This high-fidelity generation of the lowest-order maximally path-entangled state, combined with confirmation of the phase stability of the superposition, demonstrates the applicability of integrated devices for quantum metrology applications.

Finally, we tested the simple quantum circuits shown in Fig. 4, B and C, consisting of a CNOT gate and Hadamard H gates—$|0\rangle \rightarrow |0\rangle + |1\rangle$; $|1\rangle \rightarrow |0\rangle - |1\rangle$—each implemented with a 50:50 coupler between the C_0 and C_1 waveguides (25). In both cases, we observe good agreement with the ideal operation, as quantified by the average classical fidelity between probability distributions (36, 37): 97.9 ± 0.4% and 91.5 ± 0.2%, respectively. The device shown in Fig. 4B should produce equal superpositions of the four computation basis states $|00\rangle \pm |01\rangle \pm |10\rangle \pm |11\rangle$ and that shown in Fig. 4C should produce the four maximally entangled Bell states $\Psi^\pm = |01\rangle \pm |10\rangle$ and $\Phi^\pm = |00\rangle \pm |11\rangle$. Although this cannot be confirmed directly on-chip, the above demonstrations of excellent logical basis operation of the CNOT and coherent quantum operation give us great confidence.
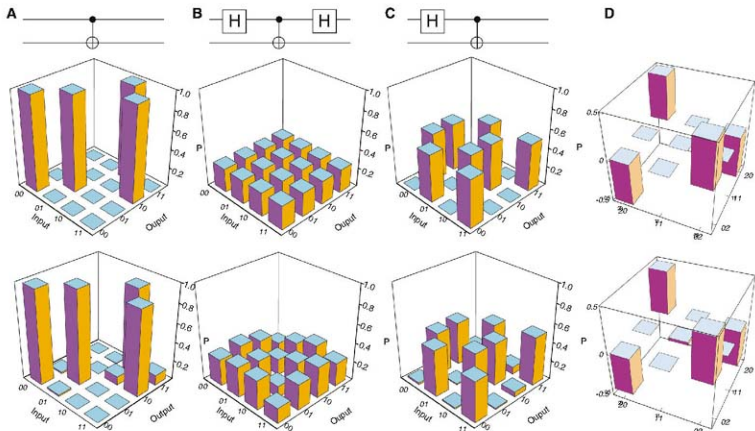
Fig. 4. Characterization of integrated quantum photonic circuits. Ideal and measured truth tables for a CNOT circuit (**A**); a CNOT with two additional H gates (**B**); and a CNOT with one additional H gate (**C**). The physical implementation fabricated on the chip is shown in fig. S1. (**D**) The ideal and estimated density matrix for the maximally path-entangled state ($|20\rangle - |02\rangle)/\sqrt{2}$.

Previous bulk optical implementations of similar photonic quantum circuits have required the design and implementation of sophisticated interferometers. Constructing such interferometers has been a major obstacle to the realization of photonic quantum circuits. The results presented here show that this problem can be drastically reduced by using waveguide devices: It becomes possible to directly write the theoretical "blackboard sketch" onto the chip, without requiring sophisticated interferometers.

We have demonstrated high-fidelity integrated implementations of each of the key components of photonic quantum circuits, as well as several small-scale circuits. This opens the way for miniaturizing, scaling, and improving the performance of photonic quantum circuits for both future quantum technologies and the next generation of fundamental quantum optics studies in the laboratory.

References and Notes
1. M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
2. N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, *Rev. Mod. Phys.* **74**, 145 (2002).
3. T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, S. Takeuchi, *Science* **316**, 726 (2007).
4. B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, G. J. Pryde, *Nature* **450**, 393 (2007).
5. J. L. O'Brien, *Science* **318**, 1393 (2007).
6. Y. Kawabe, H. Fujiwara, R. Okamoto, K. Sasaki, S. Takeuchi, *Opt. Express* **15**, 14244 (2007).
7. J. L. O'Brien, *Science* **318**, 1567 (2007).
8. P. Grangier, B. Sanders, J. Vuckovic, Eds., Special Issue: Focus on Single Photons on Demand, *New J. Phys.* **6** (2004).
9. A. Migdal, J. Dowling, Eds. Special Issue: Single-Photon: Detectors, Applications, and Measurement Methods, *J. Mod. Opt.* **51** (2004).
10. M. W. Mitchell, J. S. Lundeen, A. M. Steinberg, *Nature* **429**, 161 (2004).
11. P. Walther *et al.*, *Nature* **429**, 158 (2004).
12. K. J. Resch *et al.*, *Phys. Rev. Lett.* **98**, 223601 (2007).
13. F. W. Sun, B. H. Liu, Y. F. Huang, Z. Y. Ou, G. C. Guo, *Phys. Rev. A* **74**, 033812 (2006).
14. T. B. Pittman, M. J. Fitch, B. C. Jacobs, J. D. Franson, *Phys. Rev. A* **68**, 032316 (2003).
15. J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, D. Branning, *Nature* **426**, 264 (2003).
16. J. L. O'Brien *et al.*, *Phys. Rev. Lett.* **93**, 080502 (2004).
17. N. K. Langford *et al.*, *Phys. Rev. Lett.* **95**, 210504 (2005).
18. N. Kiesel, C. Schmid, U. Weber, R. Ursin, H. Weinfurter, *Phys. Rev. Lett.* **95**, 210505 (2005).
19. R. Okamoto, H. F. Hofmann, S. Takeuchi, K. Sasaki, *Phys. Rev. Lett.* **95**, 210506 (2005).
20. S. Gasparoni, J.-W. Pan, P. Walther, T. Rudolph, A. Zeilinger, *Phys. Rev. Lett.* **93**, 020504 (2004).
21. P. Walther *et al.*, *Nature* **434**, 169 (2005).
22. N. Kiesel *et al.*, *Phys. Rev. Lett.* **95**, 210502 (2005).
23. C.-Y. Lu *et al.*, *Nat. Phys.* **3**, 91 (2007).
24. C. K. Hong, Z. Y. Ou, L. Mandel, *Phys. Rev. Lett.* **59**, 2044 (1987).
25. Materials and methods are available as supporting material on *Science* Online.
26. All devices were fabricated at the Centre for Integrated Photonics, CIP Technologies, www.ciphotonics.com.
27. In a separate experiment with a bulk optics BS, we used this source to observe quantum interference with V = 97%.
28. Minimal effort was made to match the waveguide and fiber modes—no tapers were used, for example—and the coupling efficiency was likely limited by a mismatch of mode size and shape.
29. V = V_{min} = V_{max}.
30. P. P. Rohde, G. J. Pryde, J. L. O'Brien, T. C. Ralph, *Phys. Rev. A* **72**, 032306 (2005).
31. The waveguides are designed such that the cut-off wavelength for higher-order modes is very near to the design wavelength in order to maintain a large waveguide core size.
32. E. Knill, R. Laflamme, G. J. Milburn, *Nature* **409**, 46 (2001).
33. T. C. Ralph, N. K. Langford, T. B. Bell, A. G. White, *Phys. Rev. A* **65**, 062324 (2002).
34. H. F. Hofmann, S. Takeuchi, *Phys. Rev. A* **66**, 024308 (2002).
35. We cannot measure the four "zero" coherences in the density matrix (although they are limited by the small |11⟩⟨11| population) or distinguish between non-maximal coherences and rotated coherences with imaginary components for the |20⟩⟨02| and |02⟩⟨20| terms. However, neither of these effects changes the state fidelity. We have assumed a worst-case scenario throughout, i.e., we assume that the |11⟩ component inside the interferometer makes no contribution to two-photon detection in T_0 or T_1.
36. G. J. Pryde, J. L. O'Brien, A. G. White, S. D. Bartlett, T. C. Ralph, *Phys. Rev. Lett.* **92**, 190402 (2004).
37. T. C. Ralph, S. D. Bartlett, J. L. O'Brien, G. J. Pryde, H. M. Wiseman, *Phys. Rev. A* **73**, 012113 (2006).
38. We thank A. Clark, J. Fulconis, A. Gilchrist, A. Laing, S. Lardenois, G. Maxwell, and A. Stefanov for helpful discussions. This work was supported by the U.S. Disruptive Technologies Office (DTO), the U.K. Engineering and Physical Sciences Research Council (EPSRC), the U.K. Quantum Information Processing Interdisciplinary Collaboration (QIP IRC), and the Leverhulme Trust.

Practical Synthesis of Prostratin, DPP, and Their Analogs, Adjuvant Leads Against Latent HIV

Paul A. Wender,* Jung-Min Kee, Jeffrey M. Warrington

Although antiretroviral therapies have been effective in decreasing active viral loads in AIDS patients, the persistence of latent viral reservoirs prevents eradication of the virus. Prostratin and DPP (12-deoxyphorbol-13-phenylacetate) activate the latent virus and thus represent promising adjuvants for antiviral therapy. Their limited supply and the challenges of accessing related structures have, however, impeded therapeutic development and the search for clinically superior analogs. Here we report a practical synthesis of prostratin and DPP starting from phorbol or crotophorbolone, agents readily available from renewable sources, including a biodiesel candidate. This synthesis reliably supplies gram quantities of the therapeutically promising natural products, hitherto available only in low and variable amounts from natural sources, and opens access to a variety of new analogs.

A IDS is a pandemic disease caused by HIV. In a recent report, the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated that 33.2 million people were living with HIV and that 2.1 million people lost their lives to AIDS in the year 2007 (*1*).

Highly active antiretroviral therapy (HAART) has been successful in reducing HIV-1 levels in the plasma of many treated patients to undetectable levels. However, latent virus reservoirs remain in patients even after HAART (*2*). Such reservoirs are not targeted by current drug treatments, and as a consequence viral rebound often occurs if therapy is interrupted.

These latent viral reservoirs decrease only slowly in patients undergoing HAART. It is estimated that decades of treatment would be required to completely eliminate the latent virus. Such chronic treatment is undesirabl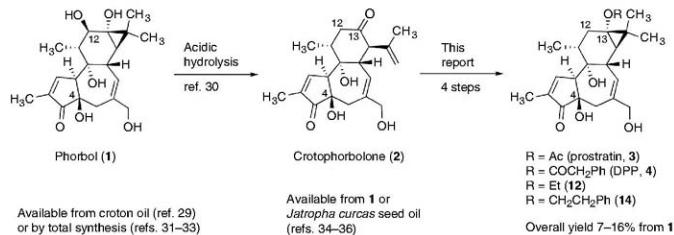e because of the increased risk of side effects over time; the emergence of resistance through viral mutation; the increased demand on patients to maintain a long-term treatment regimen; and the cumulative financial burden of prolonged therapy, a particularly problematic issue in less-developed countries. Therefore, agents that can controllably flush the latent virus from its reservoirs could, in principle, provide a means to eradicate the virus when used as adjuvants in combination with HAART (*3*).

Although agents such as interleukin-2 and valproic acid have been tested as adjuvants in HAART, they cause toxicity or efficacy problems (*4*). Phorbol-12-myristate-12-acetate (PMA), a phorbol diester, is also reported to induce HIV-1 activation, but its potent tumor-promoting activity raises concerns about its therapeutic use (*5*, *6*).

Prostratin (**3**, 12-deoxyphorbol-13-acetate) and DPP (**4**, 12-deoxyphorbol-13-phenylacetate) are non–tumor-promoting 12-deoxytigliane diterpenes that exhibit potent in vitro activity in inducing HIV expression in latently infected cell lines and primary cells (*7–11*). Prostratin and DPP also inhibit HIV entry into target cells by down-regulating CD4 and CXCR4 receptors (*12–14*).

Department of Chemistry and Department of Chemical and Systems Biology, Stanford University, 337 Campus Drive, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: wenderp@stanford.edu

Scheme 1.

(a) thiocarbonyl diimidazole (3 equiv.), DMAP (3 equiv.), CH₂Cl₂, 25 °C, 9 h, 87%;
(b) AIBN (1 equiv.), Bu₃SnH (10 equiv.), PhCH₃, 90 °C, 15 min, 75% (α:β = 1:2)
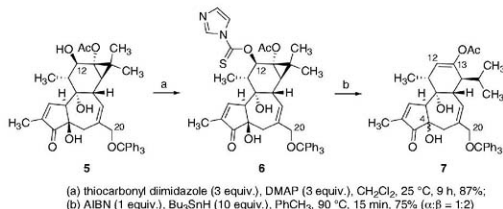
Scheme 2.

The mechanism of action of prostratin has not yet been completely elucidated, but the activation of protein kinase C (PKC) and nuclear factor κB (NF-κB) by prostratin have been proposed as key events (15–17). Prostratin has been advanced into preclinical development (18).

Unfortunately, a major obstacle to further development of prostratin, DPP, or related analogs as therapeutic agents has been their limited availability (19). Prostratin itself was first isolated from *Pimelea prostrata* and characterized by Hecker in 1976 (20). It was subsequently isolated from *Euphorbia cornigera* by Evans (21). The levels of prostratin in these source plants have, however, not been reported. More recently, impressive and seminal work by Cox in collaboration with Samoan healers and scientists at the U.S. National Institutes of Health identified prostratin as the active constituent in a traditional Samoan medicinal regimen (22). However, the Samoan source plant, *Homalanthus nutans*, affords prostratin only in low and variable isolated yields (0.2 to 52 μg/g by weight from the tree stem) (23, 24). It is noteworthy for projected clinical use of prostratin that these plant extracts do not produce acute side effects in humans, having already been used by the Samoan healers to treat individuals with certain (non-AIDS) medical conditions such as hepatitis.

The limited and varied availability of prostratin has hampered studies on its mode of action and

the identification of clinically superior analogs. Only a handful of C12-deoxytiglianes have been investigated for HIV induction activity (7, 11), and little is known about the structural basis of prostratin's biological activity. Although efforts to address this supply problem through microbial engineering have been initiated (26), the only known source of prostratin at present is plants.

Here we report practical and step economical syntheses of prostratin and DPP that can deliver research quantities (hundreds of milligrams to grams) of the targets and can serve as viable routes for addressing potential clinical needs (Scheme 1). We also show that this synthesis is sufficiently flexible to access new analogs, thereby providing an opportunity to comprehensively investigate this therapeutic lead and identify superior candidates. The step economy and flexibility of these syntheses allow facile and general access to a variety of 12-deoxytiglianes. The impact of this study may reach beyond the realm of anti-HIV therapy, because non–tumor-promoting PKC modulators such as bryostatin and its analogs represent candidates for the treatment of cancer and Alzheimer's disease (25, 27). Prostratin also inhibits the tumor-promoting effects of PMA on mouse skin (28).

Our synthesis begins with the acidic hydrolysis of phorbol (1, Scheme 1), a tigliane diterpene isolated from croton oil (29), to produce

crotophorbolone (2) (30). Croton oil is obtained from the seeds of *Croton tiglium*, a renewable source, and it is readily available in kilogram quantities in commercially long-established procedures. Phorbol itself is also commercially available, and its total synthesis has been accomplished in our laboratory (31–33). Alternatively, crotophorbolone could be obtained by the hydrolysis of 12-deoxy-16-hydroxyphorbol esters (34, 35), which are available from *Jatropha curcas* seed oil, an abundant renewable feedstock being developed as biodiesel (36, 37).

Although the synthetic conversion of phorbol to 12-deoxyphorbol (desacetylprostratin) involves the seemingly simple removal of the C12 oxygen, the proximity of this group to a strained cyclopropane ring and the sensitivity of the molecule to acid, base, heat, air, and light makes a selective deoxygenation difficult. For example, attempts to deoxygenate the C12 position of alcohol 5, derived in two steps from phorbol, by conversion to and subsequent reduction of xanthate 6 provided only enol acetate 7, arising from cleavage of the cyclopropane ring initiated by a C12 radical intermediate (Scheme 2). This was largely an anticipated result, given the known reactivity of cyclopropyl methyl radicals to undergo ring cleavage at near diffusion-controlled rates (10⁸ s⁻¹) (38). Thus, intramolecular ring cleavage occurs faster than the intermediate C12 radical can be trapped intermolecularly by an H-atom donor, despite using 10 equivalents of Bu₃SnH. Intramolecular H-atom abstraction could potentially be used to outcompete the ring opening, but even if successful, this approach would suffer from the required use of additional synthetic steps to protect the complex array of oxygens in these molecules. Like radical-based deoxygenation procedures, reductions involving conversion of the C12 hydroxyl into a good leaving group are also known to fragment the cyclopropane ring (39).

We therefore adopted a C12 deoxygenation strategy that involves first cleaving and then reestablishing the cyclopropane ring. This approach favorably moves our synthetic starting point to readily available and unprotected phorbol itself or to phorbol-derived crotophorbolone (2). With
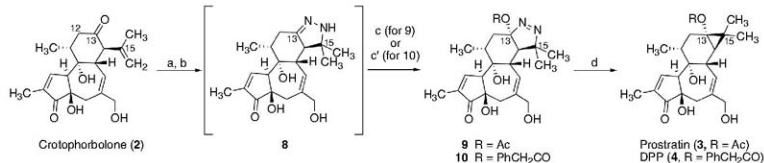
(a) $H_4N_2 \cdot H_2O$ (2 equiv.), AcOH (5 equiv.), MeOH, 25 °C, 45 min; (b) pyridine/DIPEA (9:1), 150 °C, 48 h; (c) Pb(OAc)$_4$ (1.1 equiv.), CH$_2$Cl$_2$, 0 °C, 30 min (43% of **9** from **2**); (c') Pb(OAc)$_4$ (1.2 equiv.), PhCH$_2$COOH (50 equiv.) (premixed), CH$_2$Cl$_2$, 0 °C, 30 min. (36% of **10** from **2**); (d) $h\nu$ (300 nm), EtOAc/benzene (1:1) or MeOH, 25 °C (67–92% for **3**, 90% for **4**)

Scheme 3.
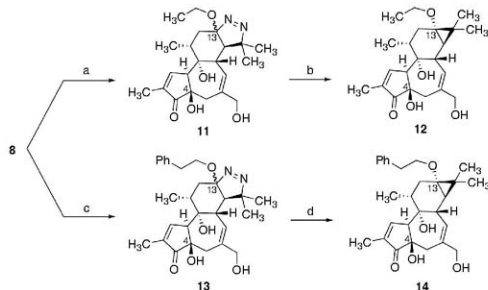
(a) PhI(OAc)$_2$ (1.2 equiv.), EtOH, 0 °C, 30 min, 24% (3:2 mixture of diastereomers, 3 steps from **2**); (b) $h\nu$ (300 nm), EtOAc, 25 °C, 72–90%; (c) PhI(OAc)$_2$ (5 equiv.), PhCH$_2$CH$_2$OH, 0 °C to 25 °C, 4 h, 18% (2:1 mixture of diastereomers, 3 steps from **2**); (d) $h\nu$ (300 nm), EtOAc, 25 °C, 87–91%.

Scheme 4.

the C12 oxygen eliminated in the formation of crotophorbolone, we reestablished the cyclopropane ring in four steps. (40) First, treatment of crotophorbolone with hydrazine in the presence of acetic acid selectively affords the C13 hydrazone (not shown in Scheme 3), which without isolation is cyclized to pyrazoline **8** when heated in the presence of base (Scheme 3). Oxidation of pyrazoline **8** with lead (IV) tetraacetate gives cyclic diazene **9**, allowing for concomitant direct introduction of a C13 acetate group and a diazene bridge between C13 and C15. Other C13 esters can also be directly introduced with this procedure by using the corresponding lead (IV) carboxylate or related oxidants. Photolysis of cyclic diazene **9** results in the extrusion of nitrogen and reestablishment of the C13-C15 cyclopropane bond (41, 42), providing prostratin (**3**) in high yield and in a remarkably concise

four-step sequence from **2** (or five steps in 12 to 16% overall yield from **1**, producing over 100 mg of **3** in a single run). The synthetic sample of prostratin so obtained was identical in all standard analytical tests to a sample of natural prostratin. This sequence can be readily conducted on a gram scale and would allow for the production of larger quantities in a proper scaleup facility (43). Moreover, this procedure is shorter than the originally considered direct C12 deoxygenation strategy (Scheme 2) because it avoids the need to introduce or remove protecting groups in these densely functionalized molecules.

An additionally attractive aspect of this synthetic strategy is that it allows access to a wide variety of analogs as exemplified below. For example, the lead (IV)-mediated pyrazoline oxidation can be easily modified to accommodate the direct introduction of other esters at the C13

position. Toward this end, when the acetate ligands of lead tetraacetate are exchanged by premixing with an excess of phenylacetic acid (44), the resulting salt induces the oxidative conversion of pyrazoline **8** to diazene **10** (in 36% yield for three steps from **2**). Subsequent photolysis affords the natural product and therapeutic lead DPP (**4**) in 90% yield, or 13% overall yield from **1**.

We have also found that this procedure can be used to access previously unknown ether analogs of prostratin (Scheme 4). Ether analogs **12** and **14** were selected as initial targets on the basis of their structural similarity to prostratin and DPP. They would also be more stable against hydrolytic decomposition than prostratin itself, which contains a hydrolytically labile ester group. The use of a seemingly straightforward, classical Williamson etherification to make such ethers from a C13 alcohol would be complicated because of the well-known facile epimerization of the C4 center in phorbol derivatives under mildly basic reaction conditions (45). Selective etherification of the C13 alcohol would also require extensive use of protecting groups to suppress reaction of the other alcohol functionalities. Instead, the ethyl ether analog **12** was readily prepared from pyrazoline **8** by treatment with PhI(OAc)$_2$ in ethanol to afford diazene **11**, which were subsequently photolyzed. Intermediate **11** is a 3:2 mixture of C13 epimers, but photolysis of either isomer gives the desired product **12** in up to 90% yield. The phenylethyl ether analog **14** was analogously synthesized in a straightforward fashion from **8** via intermediate **13**. This procedure thus provides a facile route to non-natural analogs of prostratin, based on a simple variation in our original plan.

This practical synthesis of prostratin, DPP, and other 12-deoxyphorbol esters enables the fuller investigation of their mode of action and the identification of potentially superior clinical candidates that could be used in the treatment of HIV.

References and Notes
1. UNAIDS/WHO, AIDS Epidemic Update: December 2007, available at http://data.unaids.org/pub/EPISlides/2007/2007_epiupdate_en.pdf.

2. T.-W. Chun et al., Proc. Natl. Acad. Sci. U.S.A. 94, 13193 (1997).
3. Q. E. Yang, Med. Sci. Monit. 10, 155 (2004).
4. J. D. Siliciano et al., J. Infect. Dis. 195, 833 (2007).
5. J. Laurence, S. K. Sikder, S. Jhaveri, J. E. Salmon, Biochem. Biophys. Res. Commun. 166, 349 (1990).
6. K. A. Roebuck, D. S. Gu, M. F. Kagnoff, AIDS 10, 819 (1996).
7. S. Bocklandt, P. M. Blumberg, D. H. Hamer, Antiviral Res. 59, 89 (2003).
8. A. Biancotto et al., J. Virol. 78, 10507 (2004).
9. J. Kulkosky et al., Blood 98, 3006 (2001).
10. Y. D. Korin, D. G. Brooks, S. Brown, A. Korotzer, J. A. Zack, J. Virol. 76, 8118 (2002).
11. J. Kulkosky et al., AIDS Res. Hum. Retroviruses 20, 497 (2004).
12. R. J. Gulakowski, J. B. McMahon, R. W. Buckheit Jr., K. R. Gustafson, M. R. Boyd, Antiviral Res. 33, 87 (1997).
13. M. Witvrouw et al., Antiviral Chem. Chemother. 14, 321 (2003).
14. J. Rullas et al., Antiviral Ther. 9, 545 (2004).
15. M. Hezareh et al., Antiviral Chem. Chemother. 15, 207 (2004).
16. S. A. Trushin et al., J. Virol. 79, 9821 (2005).
17. S. A. Williams et al., J. Biol. Chem. 279, 42008 (2004).
18. S. J. Brown et al., paper presented at the 15th International AIDS Conference, Bangkok, Thailand, 11 to 16 June 2004 (abstract no. TuPe B4490), available at http://gateway.nlm.nih.gov/MeetingAbstracts/102282312.html.

19. F. J. Evans, R. G. Schmidt, Acta Pharmacol. Toxicol. (Copenhagen) 45, 181 (1979).
20. A. R. Cashmore et al., Tetrahedron Lett. 17, 1737 (1976).
21. G. A. Miana, M. Bashir, F. J. Evans, Planta Med. 51, 353 (1985).
22. P. A. Cox, Pharm. Biol. 39, 33 (2001).
23. K. R. Gustafson et al., J. Med. Chem. 35, 1978 (1992).
24. H. Johnson, S. A. Banack, P. A. Cox, paper presented at the 48th Annual Meeting of Society for Economic Botany, Chicago, IL, 4 to 7 June 2007, available at www.econbot.org/_organization_/07_annual_meetings/meeting_abstracts/2007.php.
25. J. Honngaisan, D. J. Alkon, Proc. Natl. Acad. Sci. U.S.A. 104, 19571 (2007), and references therein.
26. H. Black, Scientist 18, 59 (2004).
27. P. A. Wender et al., Curr. Drug Discov. Technol. 1, 1 (2004).
28. Z. Szallasi, L. Krsmanovic, P. M. Blumberg, Cancer Res. 53, 2507 (1993).
29. D. A. Cairns, S. S. Mirvish, L. Wallcave, D. L. Nagel, J. W. Smith, Cancer Lett. 14, 85 (1981).
30. H. W. Thielmann, E. Hecker, Liebigs Ann. Chem. 728, 158 (1969).
31. P. A. Wender et al., J. Am. Chem. Soc. 111, 8957 (1989).
32. P. A. Wender, F. E. McDonald, J. Am. Chem. Soc. 112, 4956 (1990).
33. P. A. Wender, K. D. Rice, M. E. Schnute, J. Am. Chem. Soc. 119, 7897 (1997).
34. M. Gschwendt, E. Hecker, Tetrahedron Lett. 11, 567 (1970).

35. M. Hirota et al., Cancer Res. 48, 5800 (1988).
36. W. Haas, H. Sterk, M. Mittelbach, J. Nat. Prod. 65, 1434 (2002).
37. D. Fairless, Nature 449, 652 (2007).
38. V. W. Bowry, K. U. Ingold, J. Am. Chem. Soc. 114, 4992 (1992).
39. H. Bartsch, E. Hecker, Z. Naturforsch. Teil B 24, 91 (1969).
40. Materials and methods are detailed in supporting material available on Science Online.
41. P. S. Engel, L. Shen, Can. J. Chem. 52, 4040 (1974).
42. J. P. Freeman, J. Org. Chem. 29, 1379 (1964).
43. For an example of demonstrated technology for large-scale organic photolysis, see (46).
44. L. F. H. Boston, T. D. W. Claridge, M. G. Moloney, J. Chem. Soc. Perkin Trans. 1 1995, 639 (1995).
45. S.-S. Tseng, B. L. Van Duuren, J. J. Solomon, J. Org. Chem. 42, 3645 (1977).
46. B. D. A. Hook et al., J. Org. Chem. 70, 7558 (2005).
47. We thank T. Benvegnu and T. Storz-Eckerlin (Stanford University) for exploratory studies on this project. This work was supported by grants from NIH to P.A.W. (CA31841 and CA31845).

Supporting Online Material
www.sciencemag.org/cgi/content/full/320/5876/649/DC1
Materials and Methods

28 December 2007; accepted 12 March 2008
10.1126/science.1154690

Marine Polyphosphate: A Key Player in Geologic Phosphorus Sequestration

Julia Diaz,[1] Ellery Ingall,[1]* Claudia Benitez-Nelson,[2] David Paterson,[3]† Martin D. de Jonge,[3]† Ian McNulty,[3] Jay A. Brandes[4]

The in situ or authigenic formation of calcium phosphate minerals in marine sediments is a major sink for the vital nutrient phosphorus. However, because typical sediment chemistry is not kinetically conducive to the precipitation of these minerals, the mechanism behind their formation has remained a fundamental mystery. Here, we present evidence from high-sensitivity x-ray and electrodialysis techniques to describe a mechanism by which abundant diatom-derived polyphosphates play a critical role in the formation of calcium phosphate minerals in marine sediments. This mechanism can explain the puzzlingly dispersed distribution of calcium phosphate minerals observed in marine sediments worldwide.

Phosphorus is a vital macronutrient that profoundly influences global oceanic primary production on both modern and geologic time scales (1, 2). Over the past several decades, the residence time of phosphorus in the ocean has been repeatedly revised downwards as previously unidentified sedimentary sinks have been discovered (1, 3). Among these sinks are ubiquitous fine-grained authigenic apatite minerals (4), whose origin is enigmatic (5). Given the strong influence of this mineral sink on the global cycling of phosphorus and its potential impact on long-term nu-

trient availability and biological production, an understanding of the underlying mechanisms that lead to the formation and burial of apatite in modern and ancient sediments is critically important. Here, we show that polyphosphate is a key component in the formation of apatite in marine sediments.

Polyphosphate is a relatively understudied component of the marine phosphorus cycle. A linear polymer of orthophosphate units linked by phosphoanhydride bonds (fig. S1), polyphosphate is present in cells as dense, calcium-associated cytoplasmic inclusions (6). Under phosphate-enriched conditions, cultured marine algae synthesize polyphosphate as a luxury nutrient reserve (7–12). The biological synthesis of substantial amounts of polyphosphate in natural marine systems, in contrast, has been hypothesized to be inconsequential (12), as phosphorus is present at biologically limiting concentrations in much of the global ocean (1, 3). Correspondingly, investigations into the composition of marine biogenic phosphorus compounds have typically focused

on organic forms (1, 13). The lack of commonly used analytical techniques that cleanly evaluate polyphosphate within samples has further resulted in a paucity of research on the importance of this phase. With the recent development of high-resolution x-ray spectromicroscopy methods, various particulate organic, mineral, and polymeric phosphorus-containing phases like polyphosphate can now be identified and mapped at submicrometer scales. In addition, a new combined electrodialysis/reverse osmosis technique allows for a more comprehensive examination of phosphorus composition in the dissolved phase. We have developed insights into the origin and transformation of marine polyphosphate through the application of these high-resolution x-ray (14) and high-recovery electrodialysis (15, 16) techniques.

We collected organisms, sediments, and dissolved and particulate matter during April and July 2007 from Effingham Inlet, a Pacific fjord located on Vancouver Island, British Columbia (fig. S2) (16). During the spring bloom of April 2007, intracellular polyphosphate inclusions were observed in individual diatoms, including the globally ubiquitous and abundant Skeletonema spp. (fig. S3). On the basis of bulk ^{31}P– nuclear magnetic resonance (NMR) characterization of the spring bloom plankton community (16), inorganic polyphosphate represented a substantial 7% of total phosphorus in surface water biomass. Surface water dissolved phosphate concentrations were 0.5 μM, which reflects a level of phosphorus availability typical of coastal marine systems. Nutrient ratios were also consistent with phosphorus limitation in our field site (nitrogen:phosphorus = ~40). By comparison, in laboratory cultures with enriched, ~μM phosphate concentrations, Skeletonema spp. and Thalassiosira spp. can accumulate polyphosphate to correspondingly

[1]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332-0340, USA. [2]Marine Science Program and Department of Geological Sciences, University of South Carolina, Columbia, SC 29208, USA. [3]Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA. [4]Skidaway Institute of Oceanography, 10 Ocean Science Circle, Savannah, GA 31411, USA.

*To whom correspondence should be addressed. E-mail: ingall@eas.gatech.edu
†Present address: Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168, Australia.

high levels of 30% and 19 to 43% of total cellular phosphorus, respectively (10, 12). The population of diatom-dominated plankton in Effingham Inlet exhibited near-Redfield elemental stoichiometry, with a molar C:N:P composition of 188:16:1. The presence of substantial polyphosphate in these organisms thus appears not to alter their elemental content relative to the classic composition of marine phytoplankton. This finding suggests that inorganic polyphosphate has conventionally been quantified as a component of organic biomass and is not inconsistent with Redfield stoichiometry.

Polyphosphate can exist within a range of sizes and molecular weights inside cells, depending on the length of the polyphosphate polymer. Destruction of polyphosphate-containing diatoms by zooplankton grazing, viral infection, and senescence may liberate the intracellular contents of these cells, including variably sized polyphosphate inclusions. Consistent with these processes, we observed a substantial amount of polyphosphate in the <0.45-μm fraction of dissolved matter (16). In subsurface seawater samples processed by the high-recovery electrodialysis/reverse osmosis technique (15), polyphosphate accounted for ~11% of the total dissolved phosphorus pool. Previous dissolved matter characterizations do not report polyphosphate (17–19), probably because of the lower-recovery methods used in these studies. By adding the critical step of deionizing seawater samples before concentrating dissolved molecules, the combined electrodialysis/reverse osmosis technique can isolate up to 90% of marine dissolved matter, the highest recovery yet possible (15).

In addition to dissolved matter, polyphosphate was also present in sinking particles, representing 7% of total phosphorus in sinking material (16).

Individual diatoms containing intracellular polyphosphates were observed throughout the water column, which suggests that sinking polyphosphate reaches the sediment protectively encased within intact cells. In fact, diatoms have been shown to play major roles in the mineral-ballasted transport of material to depth (20, 21). Table 1 summarizes our results on the polyphosphate, total phosphorus, and biogenic silica composition of major phosphorus pools investigated in this study. Mass balance estimates based on these data demonstrate that plankton-derived polyphosphate can account for the entire polyphosphate content of sinking particles. To make this calculation, the concentration of polyphosphate in plankton and sinking material can be expressed relative to biogenic silica content, which is roughly conserved between these two pools. Using the total phosphorus concentrations of plankton and sinking particles, the silica-normalized polyphosphate content of sinking material is ~45% of that in organisms. This estimate, although based on a single particle flux measurement, shows that plankton are a plausible and sufficient source for the polyphosphate found in sinking material.

Bacterial decomposition of the organic diatom frustule matrix results in rapid dissolution of the mineral shell (22) and the consequent release of polyphosphate and other cellular contents to the sediment environment. This scenario is consistent with the relatively low biogenic silica content of Effingham Inlet surface sediments (Table 1) and with microscopy results showing damaged and vacant diatom frustules in sediments. In addition, high-resolution x-ray spectromicroscopy methods (14, 23) revealed an abundance of free 0.5- to 3-μm polyphosphate granules in surface sediments (16). This size range is similar to that observed within diatoms, again suggesting a diatom source.

X-ray fluorescence data indicated that among the hundreds of phosphorus-rich particles identified in our sediment samples, ~50% were polyphosphate, with the remaining fraction composed of apatite, a common calcium phosphate mineral (Fig. 1). Previous ^{31}P-NMR analysis of Effingham Inlet surface sediments has shown that polyphosphate accounts for 8% of the total phosphorus in surface sediment samples (24). In other studies, polyphosphate has eluded detection by bulk techniques such as ^{31}P-NMR because such methods are relatively insensitive to the presence of less prevalent phases. Because synchrotron-based x-ray spectromicroscopy is unique in its capacity to simultaneously image and chemically characterize minimally prepared particulate samples at submicrometer resolution, this highly sensitive method is key to the direct identification of less prevalent phases in a wide variety of environments (14).

Our findings demonstrate that marine polyphosphate accounts for 7 to 11% of the phosphorus in dissolved and particulate pools (Table 1). This level of abundance is comparable to that of

Table 1. Key chemical parameters of major phosphorus pools. Polyphosphate content for each pool was measured by ^{31}P-NMR. Total phosphorus (% total P) and weight percent (wt %) biogenic silica content were determined by standard chemical techniques (16). Where available, error estimates represent measurement reproducibility on the basis of replicate analyses. Analytical errors associated with the polyphosphate measurement are ±10% of the reported value. For example, a polyphosphate measurement of 7% would have an associated error of ±0.7%. Replicate total phosphorus measurements agreed to within <5%.

Phosphorus pool	Polyphosphate (% total P)	Total P (μmol g^{-1})	Biogenic silica (wt %)
Plankton	7	123.0 ± 1.7	40.4 ± 1.4
Dissolved matter	11	1.72*	–
Sinking particles	7	59.76 ± 0.42	43.7 ± 2.3
Surface sediment	8†	44.71	12.08 ± 0.65

*Units are μM.   †Data are from (24).
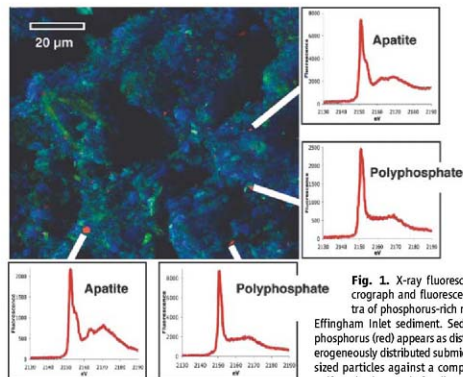
Fig. 1. X-ray fluorescence micrograph and fluorescence spectra of phosphorus-rich regions in Effingham Inlet sediment. Sedimentary phosphorus (red) appears as distinct, heterogeneously distributed submicrometer-sized particles against a comparatively uniform background of sedimentary aluminum (blue) and magnesium (green). On the basis of high-resolution x-ray spectroscopic characterization, about half of the 147 phosphorus-rich regions examined in our samples were found to be polyphosphate, whereas the other half were classified as apatite, a common calcium phosphate mineral.

more commonly identified organic phosphorus forms. For example, phosphonates typically represent ~3% of total phosphorus in fresh organic matter (25). The relative polyphosphate contents of plankton, sinking particulates, and sediments are nearly identical in our samples (Table 1). The consistency of the polyphosphate signal throughout the water column and surface sediment suggests that extracellular polyphosphate may not be readily bioavailable. Rapid enzymatic hydrolysis of polyphosphate by benthic microbes has been observed to occur, but only intracellularly (26). Because the phosphoanhydride bonds linking orthophosphate units of polyphosphate are relatively stable in the absence of hydrolytic enzymes (6), free sedimentary polyphosphate may be an efficient storage form of phosphorus over long periods. Consistent with this idea, x-ray analysis revealed the presence of extracellular polyphosphate in Effingham Inlet sediments up to 60 years old, which suggests that a portion of the free sedimentary polyphosphate pool is not remobilized over decadal time scales.

Though a portion of the sedimentary polyphosphate pool may be relatively stable, mass balance calculations reveal that some polyphosphate is removed from surface sediments over relatively short time scales. Using the bulk sediment trap flux from our field site (136 g m^{-2} year^{-1}), we estimated the sedimentary flux of polyphosphate to be 48 µg P m^{-2} day^{-1} (16). The accumulation rate of polyphosphate in recent (<3-year-old) sediments from Effingham Inlet is 86% of this flux, based on the polyphosphate and total phosphorus content of surface sediments (Table 1). This disparity reflects a 14% loss of polyphosphate in surface sediments, a loss that may increase as sediments age.

The loss of sedimentary polyphosphate in sediments does not necessarily indicate that phosphorus is remobilized from polyphosphate particles. Rather, as evidence from x-ray spectromicroscopy reveals, the relative stability of free sedimentary polyphosphate permits diagenetic transformations that result in the long-term sequestration of phosphorus. In addition to demonstrating that polyphosphate and apatite are prevalent in sediments from Effingham Inlet, results from x-ray spectromicroscopy also showed an abundance of fine, dispersed particles that exhibit spectral features transitional between pure polyphosphate and apatite (Fig. 2). Polyphosphate thus appears to nucleate authigenic apatite growth, thereby converting surface water derived polyphosphate to stable phosphorus containing mineral phases that reside in sediments over geologic time scales.

Authigenic apatite formation in marine sediments has been recognized in numerous studies as an important phosphorus sink (4). However, the processes leading to the precipitation and growth of these authigenic apatites are not well understood. Massive apatitic phosphorite deposits that account for as much as 25% of total phosphorus in the sediments underlying major

coastal upwelling zones may be related to the activity of polyphosphate-accumulating sulfur bacteria (26). Enzymatic hydrolysis of intracellular polyphosphate by these bacteria releases considerable amounts of dissolved phosphate to sediment pore waters. As a result, pore waters achieve the high degree of supersaturation required to overcome the kinetic nucleation barrier to apatite precipitation, and substantial apatite formation consequently occurs (26).

In contrast to the massive apatite-rich phosphorite formations characteristic of coastal upwelling zones, most marine sediments worldwide possess dispersed, fine-grained authigenic apatites that make up a comparatively small 9 to 13% of total sedimentary phosphorus (4). The relatively modest accumulation of authigenic apatite that is typical of sediments in nonupwelling zones nevertheless represents a substantial phosphorus sink because of the much larger areal extent of these environments (4). Authigenic apatite formation in these nonupwelling areas may not involve an episodic mechanism to produce high concentrations of dissolved phosphate, however. Rather, our results show that dispersed grains of sedimentary polyphosphate may nucleate apatite growth directly and nonepisodically, reducing or removing the nucleation barrier by acting as a mineral template. As noted previously, calcium is associated with polyphosphate in cells (6). The presence of highly concentrated sedimentary phosphorus regions with this calcium association may result in eventual apatite formation without

extensive interaction with the free sedimentary phosphate pool. We observed the transition from polyphosphate to apatite within surficial sediments <3 years of age, suggesting that apatite formation from a polyphosphate template may occur over relatively short time scales.

The transport of polyphosphate from its planktonic origin in surface waters to underlying sediments, followed by the subsequent diagenetic transformation into stable calcium phosphate minerals, provides a "biological pump" mechanism for the geologic sequestration of water column–derived marine phosphorus (fig. S4). The polyphosphate-accumulating diatoms observed in this study are common in the global ocean, including vast regions of coastal and polar seas that exhibit similar phosphate availability to our sampling site (27). Therefore, the sequestration of phosphorus through mechanisms involving diatom-derived polyphosphate is likely to be quantitatively substantial on a global scale.

Another documented source of polyphosphate in natural marine systems involves synthesis by benthic sulfur-oxidizing bacteria (26), yet these organisms thrive in specialized environments and are not as globally prevalent as diatoms. Polyphosphate has been identified in *Trichodesmium* spp. and other common marine cyanobacteria (7–9), suggesting that these organisms may be an important source of polyphosphate in the tropical and subtropical oceans where they are abundant. An abiological origin for polyphosphate is unlikely in most marine environments because abi-

Fig. 2. Diagenetic transformation of polyphosphate to apatite. An overlay of phosphorus x-ray fluorescence spectra collected from micrometer-sized phosphorus-rich regions in Effingham Inlet sediment illustrates the diagenetic transition from polyphosphate (**top**) to apatite (**bottom**). The primary phosphorus fluorescence peak occurs at 2150 eV (a). Spectral features above the primary peak reflect the local bonding environment of phosphorus. Polyphosphate, a simple linear polymer associated with calcium in cells, is characterized by a single peak 18 eV above the primary peak (b). In the diagenetic transition from polyphosphate to apatite, the association between phosphorus and calcium becomes more crystalline, which may account for the appearance of

a primary peak "shoulder" (c). As the crystalline mineral matrix develops further, a peak 11 eV above the primary peak appears (d), and secondary peaks become more defined (e). The spectra presented in this figure were collected from a single Effingham Inlet sediment sample <3 years of age. Thus, the relative ages of the particles that yielded these spectra are not known.
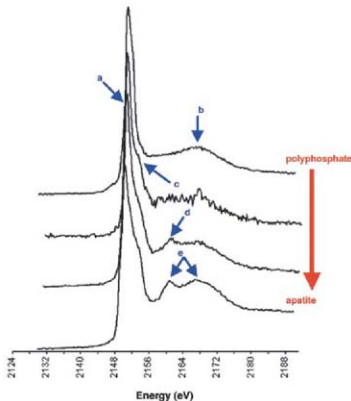
otic polyphosphate synthesis can only occur at the elevated temperatures characteristic of such extreme environments as hydrothermal vent systems (6). There is no evidence that the transformation of polyphosphate to apatite in marine sediments is dependent on the specific source of polyphosphate, however.

Enhanced phosphorus sequestration in marine sediments resulting from the conversion of diatom-derived polyphosphates to apatite may be manifested in the geologic record. The mid-Mesozoic rise of marine diatoms (28) coincides with a trend toward lower organic carbon to total phosphorus ratios in marine sediments (29). Because oceanic phosphorus influences atmospheric carbon dioxide levels over geologic time through regulation of marine primary productivity (2), geologic fluctuations in phosphorus burial efficiency brought on by changes in diatom abundance may have also exerted substantial paleo-climatic influences.

References and Notes

1. C. R. Benitez-Nelson, Earth Sci. Rev. 51, 109 (2000).
2. P. Van Cappellen, E. D. Ingall, Science 271, 493 (1996).
3. A. Paytan, K. McLaughlin, Chem. Rev. 107, 563 (2007).
4. K. C. Ruttenberg, R. A. Berner, Geochim. Cosmochim. Acta 57, 991 (1993).
5. P. Van Cappellen, R. A. Berner, Geochim. Cosmochim. Acta 55, 1219 (1991).
6. M. R. W. Brown, A. Kornberg, Proc. Natl. Acad. Sci. U.S.A. 101, 16085 (2004).
7. F. Mateo, I. Douterelo, E. Berrendero, E. Perona, J. Phycol. 42, 61 (2006).
8. K. M. Romans, E. J. Carpenter, J. Phycol. 30, 935 (1994).
9. B. A. Lawrence et al., Microbiology 169, 195 (1998).
10. L. Solorzano, J. D. H. Strickland, Limnol. Oceanogr. 13, 515 (1968).
11. K. Miyata, A. Hattori, A. Ohtsuki, Mar. Biol. 93, 291 (1986).
12. M. J. Perry, Limnol. Oceanogr. 21, 88 (1976).
13. D. M. Karl, K. M. Björkman, in Biogeochemistry of Marine Dissolved Organic Matter, D. A. Hansell, C. A. Carlson, Eds. (Academic Press, New York, 2002), chap. 6.
14. J. A. Brandes, E. Ingall, D. Paterson, Mar. Chem. 103, 250 (2007).
15. T. A. Vetter, E. M. Perdue, E. Ingall, J.-F. Koprivnjak, P. H. Pfromm, Sep. Purif. Technol. 56, 383 (2007).
16. Materials and methods are available as supporting material on Science Online.
17. L. L. Clark, E. D. Ingall, R. Benner, Am. J. Sci. 299, 724 (1999).
18. L. C. Kolowith, E. D. Ingall, R. Benner, Limnol. Oceanogr. 46, 309 (2001).
19. P. Sannigrahi, E. D. Ingall, R. Benner, Geochim. Cosmochim. Acta 70, 5868 (2006).
20. K. O. Buesseler et al., Science 316, 567 (2007).
21. R. A. Armstrong, C. Lee, J. I. Hedges, S. Honjo, S. G. Wakeham, Deep-Sea Res. II 49, 219 (2001).
22. K. Biddle, F. Azam, Nature 397, 508 (1999).
23. X-ray fluorescence spectra and phosphorus elemental maps were collected at beamline 2-ID-B, Advanced Photon Source, Argonne National Laboratory.
24. P. Sannigrahi, E. Ingall, Geochem. Trans. 6, 52 (2005).
25. L. L. Clark, E. D. Ingall, R. Benner, Nature 393, 426 (1998).
26. H. N. Schulz, H. D. Schulz, Science 307, 416 (2005).
27. F. Losurchi et al., Deep-Sea Res. I 48, 1581 (2001).
28. M. E. Katz, Z. V. Finkel, D. Grzebyk, A. H. Knoll, P. G. Falkowski, Annu. Rev. Ecol. Syst. 35, 523 (2004).
29. T. J. Algeo, E. Ingall, Palaeogeogr. Palaeoclimatol. Palaeoecol. 256, 130 (2007).
30. This material is based on work supported by NSF under grant 0526178. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences (DE-AC02-06CH11357). We thank the crew of the R/V Barnes; R. Keil and J. Nawer for assistance with field sampling; S. Herron and G. Lyons for help during field sampling and sample analysis; R. Styles and C. Jackson for assistance with sample analysis; G. Patterson and J. Matenius of the Clayoquot Field Station in Tofino, British Columbia, for providing lab space and a welcoming base for our field studies; P. Sobecky and members of her lab at Georgia Tech for the use of the epifluorescence microscope; and J. Leisen and L. Gelbaum for NMR assistance. The authors declare no competing financial interests.

Expanding Oxygen-Minimum Zones in the Tropical Oceans

Lothar Stramma,[1]* Gregory C. Johnson,[2] Janet Sprintall,[3] Volker Mohrholz[4]

Oxygen-poor waters occupy large volumes of the intermediate-depth eastern tropical oceans. Oxygen-poor conditions have far-reaching impacts on ecosystems because important mobile macroorganisms avoid or cannot survive in hypoxic zones. Climate models predict declines in oceanic dissolved oxygen produced by global warming. We constructed 50-year time series of dissolved-oxygen concentration for select tropical oceanic regions by augmenting a historical database with recent measurements. These time series reveal vertical expansion of the intermediate-depth low-oxygen zones in the eastern tropical Atlantic and the equatorial Pacific during the past 50 years. The oxygen decrease in the 300- to 700-m layer is 0.09 to 0.34 micromoles per kilogram per year. Reduced oxygen levels may have dramatic consequences for ecosystems and coastal economies.

Oceanic dissolved-oxygen concentrations affect marine biogeochemical processes and have major impacts on the global carbon and nitrogen cycles (1). These concentrations are very sensitive to changes in air-sea fluxes and interior ocean advection, hence dissolved oxygen is an important parameter for understanding the ocean's role in climate (2). Important mobile macroorganisms are stressed or die under hypoxic conditions; that is, when oxygen concentrations drop below ~60 to 120 μmol kg^{-1} (3). Hypoxia occurs at different oxygen concentrations among various species of macroorganisms, so the threshold is not precise. Regions with oxygen concentrations below about 10 μmol kg^{-1} are termed suboxic. In suboxic regions, nitrate (if present) becomes involved in respiration (1). Anoxic regions have no dissolved oxygen. At present, the intermediate-depth low-oxygen layers, here called the oxygen-minimum zone (OMZ), are suboxic in the eastern tropical Pacific Ocean and the northern reaches of the tropical Indian Ocean and are hypoxic in the tropical Atlantic Ocean (Fig. 1).

Oceanic dissolved oxygen concentrations have varied widely in the geologic past. For instance, paleoclimate records from the Cretaceous reveal profoundly altered biogeochemical cycles and dramatic consequences for ecosystems associated with reductions of ocean oxygen (4). The anoxic ocean at the end of the Permian (251 million years ago) is perhaps the most striking example, being associated with elevated atmospheric CO_2 and massive terrestrial and oceanic extinctions (5, 6).

Climate models predict an overall decline in oceanic dissolved oxygen concentration and a consequent expansion of the OMZ under global warming conditions (7), with the largest declines occurring in extratropical regions. In the tropical regions, the models predict either zonal mean oxygen increases at depths of about 200 to 1000 m in the Atlantic and Pacific Oceans (7) or moderate zonal mean oxygen decreases (8). Predicted oxygen changes in the thermocline waters result largely from solubility changes in the upstream source waters, whereas changes in the deeper waters result mainly from decreased interior advection and ongoing oxygen consumption by remineralization of sinking particulate organic matter (7).

The global ocean has warmed substantially over the past 50 years (9), and strong interannual-to-decadal variations of oxygen have been observed in the upper 100 m (10). Long-term oxygen changes have been observed and reported in the subpolar and subtropical regions (11, 12). For instance, in the subarctic Pacific at Ocean Station Papa (50°N, 145°W), declining oxygen concentrations have been reported from depths of 100 to 400 m between 1956 and 2006 (11). Ocean oxygen data from the most oxygen-poor tropical regions of the OMZ

[1]Institut für Meereswissenschaften an der Universität Kiel (IFM-GEOMAR), Düsternbrooker Weg 20, 24105 Kiel, Germany. [2]National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115, USA. [3]Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, CA 92093, USA. [4]Baltic Sea Research Institute Warnemünde, Post Office Box 301161, 18112 Rostock, Germany.

*To whom correspondence should be addressed. E-mail: lstramma@ifm-geomar.de

are limited, but some regions exist for which historical data can be augmented with data from recent survey programs to construct relatively long, quasi-continuous oxygen time series.

We constructed and analyzed oxygen time series in some select areas of the tropical oceans (Fig. 1), using quality-controlled historical data from the HydroBase 2 database (13) and more recently measured oxygen profiles. Only oxygen data collected since 1960 were used because older oxygen data are rare and the net effect of changes in the observation system on our ability to document real ocean variability is not well understood. Unfortunately, even after 1960 oxygen data in most tropical regions are too sparse to construct useful time series, because in the past most oxygen profiles were collected almost exclusively from ships dedicated to oceanographic research. Recently, a small fraction of the 3000 Argo freely drifting floats that report vertical profiles of temperature and salinity over the upper 2000 m of the ocean via satellite at 10-day intervals (14) have been equipped with oxygen sensors. These floats provide valuable oxygen profiles (15) that were used to expand our time series through 2007 in the tropical Atlantic.

Reductions in observed minimum oxygen concentrations and vertical expansion of the OMZ since 1960 are apparent in three areas of the tropical Atlantic Ocean (Fig. 2). In the oxygen-poor region of the tropical North Atlantic (Fig. 1; 10° to 14°N, 20° to 30°W), a time series of historical data was augmented with data from meridional hydrographic sections nominally along 29°W in July 2003 and 23°W in July 2006 (Fig. 2A). No seasonal signal is present in this area

(16). In the OMZ, core oxygen values decline and the OMZ expands vertically with time. The vertical extent of the layer with oxygen concentrations of <90 μmol kg^{-1} increased 85%, from a thickness of 370 m in 1960 to 690 m in 2006. In the near-equatorial Atlantic Ocean, oxygen values are higher to both the north and south (Fig. 1) because of the eastward transport of relatively oxygen-rich water within the complicated tropical current system (17). The relatively oxygen-rich water in the Atlantic Central Water originates from the poleward side of the subtropical gyre. Historical data, hydrocasts from three repeat sections along 23°W since 2000, and two recent profiles from an Argo float allowed construction of a time series (Fig. 2B) in the central equatorial Atlantic (Fig. 1; 3°S to 3°N, 18° to 28°W). This time series also shows some indication of a reduced concentration at the vertical oxygen minimum over time and a vertical expansion of the oxygen-poor OMZ. Similarly, a tropical South Atlantic (Fig. 1; 14° to 8°S, 4° to 12°E) time series (Fig. 2C) also shows a vertical expansion of the OMZ; although there are long gaps after the late 1980s, and no data between 2001 and 2007, a recent preliminary calibrated oxygen profile taken at 9°S, 8°E in March 2008 is consistent with the trend of an oxygen decline.

The OMZ in the tropical North and South Pacific Oceans reaches suboxic (and, in the most oxygen-poor regions, nearly anoxic) levels (Fig. 1), so detecting changes in minimum values there is difficult. Furthermore, data in the most oxygen-poor regions are too sparse to allow the construction of quasi-continuous time

series. However, as in the Atlantic Ocean, Pacific equatorial currents carry relatively oxygen-rich water eastward toward the most oxygen-poor regions of the OMZ in both hemispheres. Historical hydrographic data in the eastern equatorial Pacific Ocean (Fig. 1; 5°S to 5°N, 105° to 115°W), augmented with data collected during some recent Tropical Atmosphere Ocean project mooring maintenance cruises along 110°W, constitute a time series (Fig. 2D) that reveals a vertical expansion of the OMZ. However, a depth-integrated oxygen trend there is not statistically significantly different from zero when a stringent 95% confidence criterion is used (Table 1). Slightly higher values from 1980 to 1990 may be caused by sample locations biased toward the equator, where more oxygen-rich waters are advected eastward from the west. In the central equatorial Pacific (Fig. 1; 5°S to 5°N, 165° to 175°W), oxygen concentrations within the OMZ are more variable (Fig. 2E). Nevertheless, the OMZ thickness expands over the duration of the time series. This vertical expansion with time is not closely related to a temperature increase; in both areas of the equatorial Pacific, the temperature in the 300- to 700-m layer slightly decreases, as does the oxygen content (Table 1).

In the Indian Ocean, the lowest oxygen values in the OMZ are not located in the eastern tropics as they are in the Atlantic and Pacific Oceans, but to the north in the Arabian Sea and the Bay of Bengal (Fig. 1). In addition, minimum oxygen concentrations within the Indian Ocean OMZ are generally deeper (near 800 m) than in the other two oceans. In the northern Indian Ocean, sources and sinks of oxygen are nearly in apparent balance; circulation there appears relatively stagnant, with detritus falling from the highly productive waters above and rapidly depleting oxygen below. As in the eastern tropical Pacific Ocean, oxygen values in the northern Indian Ocean OMZ are suboxic, and the sparse data distributions in the most oxygen-poor regions preclude the construction of long quasi-continuous time series there. However, the recent occupation of a meridional section nominally along 95°E made possible the construction of an eastern equatorial Indian Ocean time series (Fig. 1; 5°S to 0°, 90° to 98°E), despite gaps in data since the mid-1980s (Fig. 2F). Unlike the other time series presented here, there is neither an obvious increase of the vertical extent of the OMZ nor a visible decrease in oxygen minimum values (Table 1). Time series from the early 1960s to the late 1990s (not shown) in the western equatorial Indian Ocean, the Arabian Sea, and the Bay of Bengal show similar constancy in the tropical Indian Ocean OMZ. Collectively, these results suggest that over the past few decades there has been no substantial change in the tropical Indian Ocean OMZ.
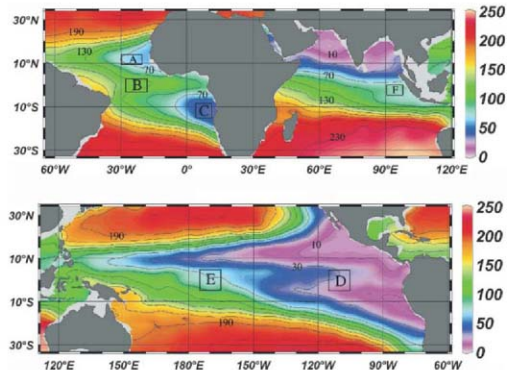
Fig. 1. Climatological mean (18) dissolved oxygen concentrations (μmol kg^{-1} shown in color) at 400 m depth contoured at 20-μmol-kg^{-1} intervals from 10 to 230 μmol kg^{-1} (black lines) using Ocean Data View (19) software. Analyzed areas (A to F, Table 1, and Fig. 2) are enclosed by black boxes.

As an auxiliary benefit, the constancy of OMZ characteristics in the Indian Ocean time series suggests that the changes observed in the tropical Atlantic and Pacific Ocean OMZ characteristics are not based on changes in observa-tion techniques. This concern was also tested in the oxygen-rich deep-water formation region of the Labrador Sea, for which a very well-sampled time series can be constructed from 1960 to the present. There, no trend toward lower oxygen val-ues is apparent, supporting the claim that oceanic oxygen measurements taken over the past 50 years are not subject to large observational biases that may produce spurious temporal trends.

The tropical ocean OMZs in the central and eastern tropical Atlantic and equatorial Pacific Oceans appear to have expanded and intensified during the past 50 years. Despite the sparseness of observations, the time series used show that the decline in oxygen content has been most intense in the tropical Atlantic, where at present hypoxic regions are small as compared with the Pacific and Indian Oceans. For these reasons, the Atlantic may also have the most potential for large increases in the area of hypoxic regions.

The observational analysis presented here sup-ports climate model predictions of dissolved oxygen declines in the tropical ocean (7, 8) and an expansion of the tropical OMZs due to a con-tribution of thermal, dynamical, and biogeochem-ical factors (8). The observed oxygen declines reported here of 0.09 to 0.34 μmol kg^{-1}year^{-1} for 300- to 700-m depths (Table 1) are somewhat smaller than those reported in the North Pacific (11) at 100 to 400 m. Together, these trends affect carbon and nitrogen cycles, with fundamental implications for marine ecosystems and thereby fisheries resource management issues. Given climate model projections, and the geological record that indicates times of widely distributed suboxic regions, sustained global ocean measure-ments of dissolved oxygen concentrations are needed (for instance, by equipping more Argo floats with well-calibrated dissolved oxygen sensors) to more closely monitor variations in the strength and extent of the OMZ.
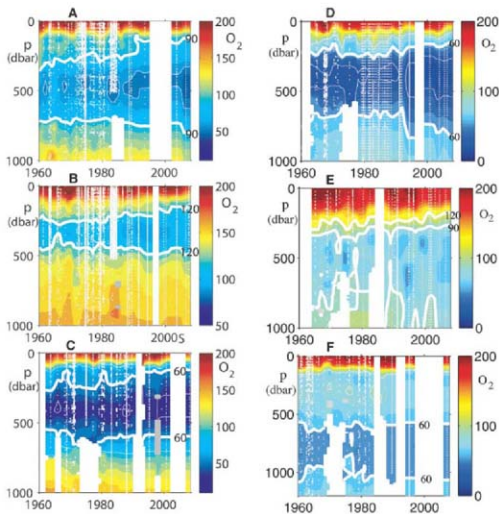
Fig. 2. Dissolved oxygen concentration (μmol kg^{-1} shown in color) maps (20, 21) versus time (1960–2008) and pressure (1 dbar ~ 1 m) with sample locations (white dots). (**A**) The eastern tropical North Atlantic (10° to 14°N, 20° to 30°W), contoured at 90 μmol kg^{-1} (thick white line). (**B**) The central equatorial Atlantic (3°S to 3°N, 28° to 18°W), contoured at 120 μmol kg^{-1} (thick white line). (**C**) The eastern tropical South Atlantic (14° to 8°S, 4° to 12°E), contoured at 60 μmol kg^{-1} (thick white line). (**D**) The eastern equatorial Pacific Ocean (5°S to 5°N, 105° to 115°W), contoured at 60 μmol kg^{-1} (thick white line). (**E**) The central equatorial Pacific Ocean (5°S to 5°N 165° to 175°W), contoured at 90 and 120 μmol kg^{-1} (thick white lines). (**F**) The eastern equatorial Indian Ocean (5°S to 0, 90° to 98°E), contoured at 60 μmol kg^{-1} (thick white line).

Table 1. Linear trends of temperature and oxygen with 95% confidence intervals (22) since 1960 in a 300- to 700-m layer for select ocean areas, and integrated oxygen loss, assuming a nominal density of 1027.2 kg m^{-3}.

Ocean areas (Fig. 1)	Temperature trend (°C year^{-1})	Oxygen trend (μmol kg^{-1} year^{-1})	Integrated oxygen loss (mmol m^2 year^{-1})
Area A	+0.009 ± 0.008	−0.34 ± 0.13	136
Area B	+0.005 ± 0.008	−0.19 ± 0.12	74
Area C	+0.002 ± 0.011	−0.17 ± 0.11	74
Area D	−0.001 ± 0.009	−0.13 ± 0.32	49
Area E	−0.010 ± 0.008	−0.19 ± 0.20	74
Area F	+0.005 ± 0.007	−0.09 ± 0.21	37
N. Pacific, 100 to 400 m depth (11)	+0.005 to +0.012	−0.39 to −0.70	165

References and Notes
1. H. W. Bange, S. W. Naqvi, L. A. Codispoti, *Prog. Oceanogr.* **65**, 145 (2005).
2. F. Joos et al., *Eos* **84**, 197 (2003).
3. J. S. Gray, R. S. Wu, Y. Y. Or, *Mar. Ecol. Prog. Ser.* **238**, 249 (2002).
4. C. E. Jones, H. C. Jenkins, *Am. J. Sci.* **301**, 112 (2001).
5. M. J. Benton, R. J. Twitchett, *Trends Ecol. Evol.* **18**, 358 (2003).
6. P. B. Wignall, R. J. Twitchett, *Science* **272**, 1155 (1996).
7. R. J. Matear, A. C. Hirst, *Global Biogeochem. Cycles* **17**, 10.1029/2002GB001997 (2003).
8. L. Bopp, C. Le Quere, M. Heimann, A. C. Manning, P. Monfray, *Global Biogeochem. Cycles* **16**, 10.1029/2001GB001445 (2002).
9. S. Levitus, J. I. Antonov, T. P. Boyer, C. Stephens, *Science* **287**, 2225 (2000).
10. H. E. Garcia, T. P. Boyer, S. Levitus, R. A. Locarnini, J. Antonov, *Geophys. Res. Lett.* **32**, 10.1029/2004GL022286 (2005).
11. F. A. Whitney, H. J. Freeland, M. Robert, *Prog. Oceanogr.* **75**, 179 (2007).
12. R. J. Matear, A. C. Hirst, B. I. McNeil, *Geochem. Geophys. Geosyst.* **1**, 10.1029/2000GC000086 (2000).
13. R. Curry, www.whoi.edu/science/PO/hydrobase/ (2007).
14. W. J. Gould, J. Turton, *Weather* **61**, 17 (2006).
15. A. Körtzinger, J. Schimanski, U. Send, D. R. W. Wallace, *Science* **306**, 1337 (2004).
16. L. Stramma et al., *J. Geophys. Res.* **113**, C04014, 10.29/2007JC004369 (2008).
17. L. Stramma, J. Fischer, P. Brandt, F. Schott, in *Interhemispheric Water Exchange in the Atlantic Ocean*, G. J. Goni, P. Malanotte-Rizzoli, Eds. (Elsevier, Amsterdam, 2003), pp. 1–22.

18. H. E. Garcia, R. A. Locarnini, T. P. Boyer, J. I. Antonov, in *World Ocean Atlas 2005, Vol. 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation,* S. Levitus, Ed. (National Oceanic and Atmospheric Administration Atlas National Environmental Satellite, Data, and Information Service 63, U.S. Government Printing Office, Washington, DC, 2006), p. 1–342.

19. R. Schlitzer, http://odv.awi.de (2007).

20. HydroBase (*13*) quality-controlled data augmented with repeat transects and available Argo float profiles within the areas shown (Fig. 1) were objectively mapped using correlation scales of 1 year and 50 m. To ameliorate potential spatial bias due to recent sparser sampling, the meridional conductivity temperature depth and bottle sections used after 2000 were longitudinally centered within each area. Inspection of data distributions

21. suggests that reported trends are not based on geographical shifts of data locations inside the investigation areas or seasonal shifts as a function of time, except as noted in the text.

21. Suspect data for 1989 were removed from Fig. 2D (offset from the surface to 1000 m), for 1963 from Fig. 2E (stations only in the southern part of the box), and for 1986 from Fig. 2F (oxygen increased from 400 m to a maximum at 1000 m).

22. Linear trends and their 95% confidence intervals were estimated as in (*23*), using annual 300- to 700-m averages of the objectively mapped fields. Degrees of freedom for the confidence intervals were determined from integral time scales as in (*24*).

23. C. Wunsch, *The Ocean Circulation Inverse Problem* (Cambridge Univ. Press, Cambridge, 1996).

24. H. von Storch, F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge Univ. Press, Cambridge, 1999).

25. The Deutsche Forschungsgemeinschaft provided support as part of the German project Sonderforschungsbereich 754 (U.S.). Additional support was provided through the National Oceanic and Atmospheric Administration (NOAA) Office of Oceanic and Atmospheric Research (G.C.J.) and NSF award no. 0223869 (J.S.). Findings and conclusions in this article are those of the authors and do not necessarily represent the views of NOAA. Float data are collected and made freely available by the international Argo Project and contributing national programs (www.argo.ucsd.edu).

6 December 2007; accepted 20 March 2008
10.1126/science.1153847

A General Model for Food Web Structure

Stefano Allesina,[1,2]* David Alonso,[1,3] Mercedes Pascual[1,4]

A central problem in ecology is determining the processes that shape the complex networks known as food webs formed by species and their feeding relationships. The topology of these networks is a major determinant of ecosystems' dynamics and is ultimately responsible for their responses to human impacts. Several simple models have been proposed for the intricate food webs observed in nature. We show that the three main models proposed so far fail to fully replicate the empirical data, and we develop a likelihood-based approach for the direct comparison of alternative models based on the full structure of the network. Results drive a new model that is able to generate all the empirical data sets and to do so with the highest likelihood.
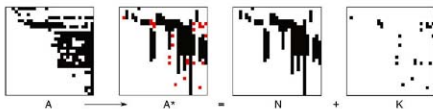
Food webs (*1–3*) are paradigmatic examples of complex systems in nature (*4*). Despite the challenge posed by the intricacy of these trophic networks, simple models have been proposed for their topology that successfully capture a number of structural properties (*5–7*). These models have been influential in showing that the topology of food webs in nature is non-random and have provided a basis for investigating the consequences of their structure for dynamics (*8, 6, 9*), for an ecosystem's robustness to extinctions (*10*), and for the quantity and quality of services they provide (*11, 12*).

The simplest mathematical framework for food web structure dates back to the influential argument on stability and complexity, and it relied on the representation of connections between species based on random graphs (*13*). This model took into account only the species richness S and connectance C (fraction of realized feeding connections) of the web. The first nonrandom representation was given by the cascade model (*5*), which ordered species along a single dimension.

The biological basis for this ordering remains an open problem, but possibilities include body mass, trophic level, and metabolic rates (*14, 9*). Each species has a position in this hierarchy that determines its feeding relations, with prey chosen randomly only from the species whose ranking is lower than that of the predator. This rule makes all networks generated by this model acyclic, limiting its application to empirical food webs without cannibalism or feeding cycles. The niche model was proposed next (*6, 15*), in part to overcome this limitation. It retains the ordering of species in one dimension but adds the notion of a niche range, an interval that contains all the prey of a given predator. Although feeding cycles can now be generated, the resulting networks are, by construction, also interval, a property that is not

fully compatible with patterns in empirical food webs (*7, 16*).

Intervality has played an important role in the literature of food web models, because it is closely related to the number of dimensions needed to represent niches in a community (*17, 7, 16*). Technically, this property means that there exists a suitable ordering of the species for which all the prey of each predator are consecutive, with no gaps. In Fig. 1, matrix N, this property is apparent when the network is translated into a matrix representation; consecutive prey form an uninterrupted sequence of entries in each column. Although, for interval graphs, a single dimension should be sufficient, recent analyses indicate that food webs are only close to interval (*16*). As we show here, close to interval does not mean that a model assuming perfect intervality on a single axis can generate all the links in empirical food webs. A third and more recent model, the "nested hierarchy" (*7*), does not rely on niches in a one-dimensional space, but focuses instead on groups of species and considers implicitly phylogenetic constraints and adaptation (*7*). Closely related predators tend to share their prey with occasional departures from this phylogenetic constraint, as the result of adaptation to new environments and new prey (*7*). We focus here on these three static models of food web structure as the simplest and most used formulations; other models have been proposed that include more sophisticated construction rules, including dynamics and diet optimization, speciation, extinction, evolution (*18–20*), and adaptation (*21*).

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Natural Science Building, 830 North University, Ann Arbor, MI 48109, USA. [2]National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, CA 93101, USA. [3]Community and Conservation Ecology Group, Center for Ecological and Evolutionary Studies, University of Groningen, Post Office Box 14, 9750 AA Haren, Netherlands. [4]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA.

*To whom correspondence should be addressed. E-mail: allesina@nceas.ucsb.edu

Fig. 1. Decomposition of a food web into two subwebs. For a given food web, we can write an adjacency matrix A. In this matrix, each coefficient represents the presence (1, black) or lack (0, white) of interaction between predator species (columns) and prey species (rows). We seek an ordering of the species that minimizes the number of irreproducible connections, the links that are incompatible with the assumptions of a given model, in this case, the niche model. This yields the adjacency matrix A^*. The compatible connections of each predator i do fall into a segment (intervality) such that the segment starts either before or on the ith species (hierarchy). The matrix N is formed by all the connections compatible with the niche model, and the matrix K contains all the irreproducible connections.

The models have been previously compared with empirical data by simulating an ensemble of networks and by measuring a large set of summary statistics, such as the number of basal, intermediate, and top species; the average path length; the fraction of cannibalistic species; the degree of omnivory; and so on (6, 22, 23), for these networks. The performance of the model is inversely proportional to the distance between the simulated values and their empirical counterparts. The comparison of models based on this approach has several limitations. First, network properties are not independent. Second, the notion of distance requires knowledge of the natural variance of the different measures. In the absence of this information, the distribution of the values themselves has been used (6). Third, some models perform well for some indices and worse for others, which makes definitive comparisons elusive (22, 23), despite the increase in the number of network properties that have been considered. Possibly, the most striking caveat of the use of summary statistics is that it cannot tell us whether or not a model is able to fully replicate empirical networks.

We propose here a different approach based on likelihoods and, therefore, on the topology of the networks as a whole. That is, on the entire set of links that specify who consumes whom in the ecosystem, rather than on a collection of summary descriptors of structure. The likelihood that a model generates the observed data provides a single quantity that allows direct comparison between alternative models. We derive the likelihood for the cascade, niche, and nested-hierarchy models [(24), sections S1 to S3] and compare them.

Our starting point is that, strictly speaking, all the models have a likelihood of zero because they all fail to reproduce a subset of the links in the empirical webs. Thus, we arrive at a major limitation of current models: they are not general. For any ordering of the species, there are links that do not fulfill their assumptions. We refer to those links as irreproducible connections. Although this limitation has been known for the cascade and niche models (6, 7) it has not been addressed

before for the nested hierarchy. We show that this model also fails to replicate the empirical data: Most of the irreproducible links involve either cannibalism [a feeding relationship underrepresented by this model (7)] or lower-level species preying on higher level species [(24), section S3].

We therefore propose an additional step that consists of decomposing the original food web into two parts, the first composed of the links that are compatible with the specific model of interest and the second of those that the model cannot reproduce. Because the number of irreproducible connections depends on the ordering of the species in the trophic hierarchy [(24), sections S1 to S4], we used a genetic algorithm to find the ordering that minimizes this number. The process is sketched in Fig. 1 for the niche model. This then allows us to effectively formulate and compute a nontrivial likelihood for empirical webs. Each food web can be represented using a matrix (Fig. 1): We split the data into two matrices we call N and K for the reproducible and irreproducible connections, respectively. We then computed the probability of obtaining the matrix N using the model of interest and the probability of obtaining K using a random graph. The product of these two numbers gives us a total likelihood for the model. Although one could devise other ways to approach the problem of computing a likelihood, different from our decomposition, this would require specific assumptions about measurement errors [(24), section S1].

The number of irreproducible connections for each model and data set, together with the total likelihood, is shown in Table 1. The niche model has better likelihoods for all the considered food webs. However, the number of irreproducible connections for the niche model is much higher than that for the cascade and nested-hierarchy models. Thus, the niche model can reproduce a smaller subset of connections very well. This provides support for the central idea of this model that predators tend to consume prey that share common characteristics. However, the large number of irreproducible links indicates that one

single dimension is not sufficient to describe the similarity between prey animals or plants, a limitation that was raised before (7, 16).

To produce a general model that is able to generate the entire network, we then start from the niche model, as it provides a better baseline likelihood, and propose a simple way to address the multidimensionality of niche ranges. Predators can choose their prey according to several traits, such as prey body mass, movement, time of the day when foraging, color, presence or absence of antipredator behavior, and so on. A simple hypothetical example of the consequences of using a single trait to recover, from food web links, the range of predators' preference is illustrated in Fig. 2. When more than one trait underlies prey choice, the use of a single trait leads both to discontinuous ranges that contain gaps and to effective ranges that are smaller than the real ones. We therefore recover from network data only a minimum range and not its actual extent [(24), section S5].

A simple way to consider multidimensional niches [(24), sections S4 to S6] is to extend the niche model by including gaps into the diets of predators. In this minimum potential niche model, species are still ordered in a one-dimensional space, but each predator chooses a potential range (Fig. 2C). The two species at the extremes of the range are always prey (Fig. 2D, in black), and therefore, they delimit the extent of the minimum potential range. The other species contained in the range will be prey with probability $(1 - f)$, where f is the probability of a "forbidden link." Forbidden links have been introduced in the study of mutualistic networks to describe plant-animal interactions that are precluded by biological constraints, such as the short tongues of certain bees unable to efficiently pollinate long-corolla flowers (25–27). In our model, forbidden links implicitly take into account the existence of traits that are not explicitly considered when species and their niches are represented using a single dimension. The value of f specifies the fraction of nonrealized feeding interactions inside a potential range.

Table 1. Likelihood of the models for food web structure. For each model, we report the number of links (L), the number of irreproducible connections (I), the log-likelihood of obtaining such connections using random graphs [$\mathcal{L}(K)$] and the total log-likelihood for the model (Tot $\mathcal{L}$). $\mathcal{L}(N)$ can be obtained by difference. The Minimum potential niche model has no irreproducible con-

nections and results in better likelihoods for all cases. For this model, we show the probability of forbidden links (f). Note that the number of parameters is the same for all models, which allows a direct comparison of the likelihoods. In other cases, one can use criteria for model selection based on likelihoods, such as Akaike's information criterion.

Food web	S	L	Cascade			Niche			Nested hierarchy			Min. potential	
			I	$\mathcal{L}(K)$	Tot $\mathcal{L}$	I	$\mathcal{L}(K)$	Tot $\mathcal{L}$	I	$\mathcal{L}(K)$	Tot $\mathcal{L}$	Tot $\mathcal{L}$	f
Benguela	29	203	12	−62.91	−343.62	23	−105.46	−234.22	1	−7.73	−349.39	−213.52	0.170
Bridge	25	107	4	−24.19	−217.16	1	−7.44	−94.42	1	−7.44	−162.32	−92.18	0.013
Broom	85	223	4	−33.99	−857.42	36	−226.77	−737.56				−626.54	0.336
Chesapeake	31	68	1	−7.87	−199.59	10	−55.60	−166.84	3	−20.30	−200.15	−145.11	0.314
Coach	29	262	41	−163.85	−443.67	37	−151.75	−296.76	7	−40.49	−381.57	−296.10	0.240
Grass	61	97	0	0	−379.31	10	−69.18	−327.08	13	−86.52	−437.81	−294.94	0.243
Reef	50	556	59	−279.34	−1106.54	196	−687.11	−970.28	22	−126.03	−1053.50	−934.71	0.416
Skip	25	197	12	−59.32	−259.02	22	−95.24	−191.11	5	−29.12	−254.74	−169.67	0.142
St. Marks	48	221	3	−22.93	−576.69	72	−320.40	−546.48	18	−105.27	−634.04	−504.49	0.554
St. Martin	42	205	0	0	−472.58	52	−234.48	−421.53	10	−61.70	−531.55	−388.06	0.443
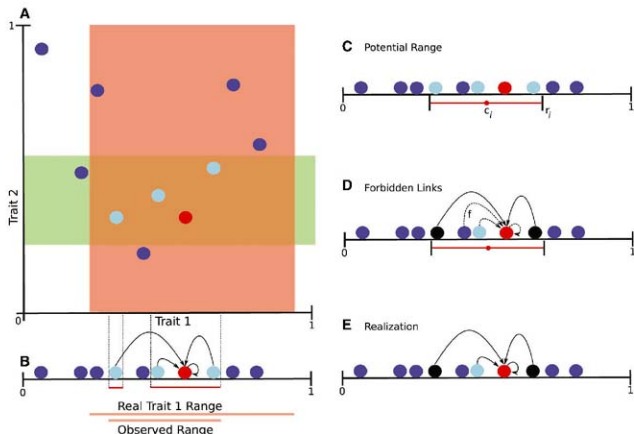
Fig. 2. Multidimensional niches and the minimum potential niche model. (**A**) A predator (in red) preys upon all the species falling in the overlap between two traits (red and cyan). If one tries to recover information on the range of trait 1 using the predator's diet, the observed range not only is smaller than the real one (**B**) but contains gaps. The minimal potential niche model starts by assigning to each species a potential range, as in the niche model (**C**). This is done by considering that the first and the last species falling in the interval are prey (**D**) (black), and that the other species falling in the interval are prey with probability $1 - f$ (dashed line in (D)) where f provides a measure on nontrivial intervality [(24), section S4]. This model can reproduce discontinuous diets (**E**).

As a consequence, our model considers an effective niche range that is a subset of a fundamental (or potential) range. A potential range, in order to include a forbidden link, must encompass at least three potential prey. Because of this, f represents a measure of nontrivial intervality in the network [(24), section S4].

The minimum potential niche model is a general model because it can reproduce all the links in the empirical food webs [e.g., (Fig. 2E)]. The derivation of the likelihood for this model follows as a straightforward modification of that for the niche model [(24), section S4]. The model requires three parameters: the number of species, S; the density of potential connections, C_p; and the probability of forbidden links, f. All three can be obtained from the empirical data. For all empirical data sets, the minimum potential niche model has the highest likelihood (Table 1).

Another simple solution to multidimensional niches and the related departure from intervality was previously proposed in a complementary approach known as the generalized niche model (GNM) (16). Instead of gaps within intervals, the GNM introduces "satellite prey" outside niche ranges. Despite the apparent similarity to the minimum potential niche model, the implementation of this approach appears problematic, and the model itself is unable to reproduce all links [(24), section S7].

In summary, the similarity between prey of a common predator on a one-dimensional niche space, which is the basis for the niche model, is well supported by empirical data. However, this property alone cannot account for all feeding relations in ecosystems, as previously noted (7, 16). We have proposed a simple way to take into account the multidimensionality of niches and to derive parameter values from the data. The proposed analysis, based on likelihoods, opens up the possibility of evaluating other fundamental assumptions of food web models, such as the existence of a hierarchy and the increased generality of predators along the niche axis. The derivation of a general model and its likelihood are also a critical first step toward evaluating the biological basis for the niche axis. It has been proposed that this axis could be mapped onto body mass, trophic levels, degree of specialization, and other characteristics of species (14, 28–30). The likelihood approach can be used for the quantitative testing of these hypotheses. Finally, this work sets a benchmark that can now be challenged by other, better models, for food web structure. The decomposition into compatible and incompatible links provides a natural starting point for improving particular models. There will always be types of models too complex for comparisons based on likelihoods, such as those that incorporate evolutionary processes explicitly

(18–21). However, simple models provide an opportunity to investigate which of the biological indices previously used for model comparisons better reflects the likelihood of a model.

References and Notes
1. S. L. Pimm, *Food Webs* (Chapman & Hall, New York, 1982).
2. J. A. Dunne, in *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford Univ. Press, Oxford, 2006), chap. 2.
3. J. M. Montoya, S. L. Pimm, R. V. Sole, *Nature* **442**, 259 (2006).
4. S. A. Levin, *Ecosystems* **1**, 431 (1998).
5. J. E. Cohen, C. M. Newman, *Proc. R. Soc. London Ser. B* **224**, 421 (1985).
6. R. J. Williams, N. D. Martinez, *Nature* **404**, 180 (2000).
7. M.-F. Cattin, L.-F. Bersier, C. Banasek-Richter, R. Baltensperger, J. P. Gabriel, *Nature* **427**, 835 (2004).
8. S. L. Pimm, J. H. Lawton, J. E. Cohen, *Nature* **350**, 669 (1991).
9. U. Brose, R. J. Williams, N. D. Martinez, *Ecol. Lett.* **9**, 1228 (2006).
10. J. A. Dunne, R. J. Williams, N. D. Martinez, *Ecol. Lett.* **5**, 558 (2002).
11. D. E. Schindler, S. R. Carpenter, J. J. Cole, J. F. Kitchell, M. L. Pace, *Science* **277**, 248 (1997).
12. J. M. Montoya, M. A. Rodríguez, B. A. Hawkins, *Ecol. Lett.* **6**, 587 (2003).
13. R. M. May, *Nature* **238**, 413 (1972).
14. G. Woodward *et al.*, *Trends Ecol. Evol.* **20**, 402 (2005).
15. R. J. Williams, N. D. Martinez, *Eur. Phys. J. B* **38**, 297 (2004).
16. D. B. Stouffer, J. Camacho, L. A. N. Amaral, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19015 (2006).
17. J. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4533 (1977).
18. G. Caldarelli, P. G. Higgs, A. J. McKane, *J. Theor. Biol.* **193**, 345 (1998).

19. N. Loeuille, M. Loreau, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5761 (2005).
20. A. G. Rossberg, H. Matsuda, T. Amemiya, K. Itoh, *Ecol. Complex.* **2**, 312 (2005).
21. A. G. Rossberg, H. Matsuda, T. Amemiya, K. Itoh, *J. Theor. Biol.* **241**, 552 (2006).
22. N. D. Martinez, L. J. Cushing, in *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford Univ. Press, Oxford, 2006), Box A, p. 87.
23. L.-F. Bersier, H.-F. Cattin, C. Banasek-Richter, R. Baltensperger, J.-P. Gabriel, in *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford Univ. Press, Oxford, 2006), Box B, p. 91.
24. Details of the model are given in *Science* Online.
25. P. Jordano, J. Bascompte, J. M. Olesen, *Ecol. Lett.* **6**, 69 (2003).
26. D. P. Vazquez, *Oikos* **108**, 421 (2005).
27. L. Santamaría, M. A. Rodríguez-Gironés, *PLoS Biol.* **5**, e31 (2007).
28. M. G. Neubert, S. C. Blumenshine, T. Jonsson, B. Rashleigh, *Oecologia* **123**, 241 (2000).
29. C. A. Layman, K. O. Winemiller, D. A. Arrington, D. B. Jepsen, *Ecology* **86**, 2530 (2005).
30. T. Jonsson, J. E. Cohen, S. R. Carpenter, *Adv. Ecol. Res* **36**, 1 (2005).
31. We thank The Center for the Study of Complex Systems at the University of Michigan for computational resources.

ROS-Generating Mitochondrial DNA Mutations Can Regulate Tumor Cell Metastasis

Kaori Ishikawa,[1,2,3]* Keizo Takenaga,[4,5]* Miho Akimoto,[5] Nobuko Koshikawa,[4] Aya Yamaguchi,[1] Hirotake Imanishi,[1] Kazuto Nakada,[1,2] Yoshio Honma,[5] Jun-Ichi Hayashi[1]†

Mutations in mitochondrial DNA (mtDNA) occur at high frequency in human tumors, but whether these mutations alter tumor cell behavior has been unclear. We used cytoplasmic hybrid (cybrid) technology to replace the endogenous mtDNA in a mouse tumor cell line that was poorly metastatic with mtDNA from a cell line that was highly metastatic, and vice versa. Using assays of metastasis in mice, we found that the recipient tumor cells acquired the metastatic potential of the transferred mtDNA. The mtDNA conferring high metastatic potential contained G13997A and 13885insC mutations in the gene encoding NADH (reduced form of nicotinamide adenine dinucleotide) dehydrogenase subunit 6 (*ND6*). These mutations produced a deficiency in respiratory complex I activity and were associated with overproduction of reactive oxygen species (ROS). Pretreatment of the highly metastatic tumor cells with ROS scavengers suppressed their metastatic potential in mice. These results indicate that mtDNA mutations can contribute to tumor progression by enhancing the metastatic potential of tumor cells.

Because most chemical carcinogens bind preferentially to mitochondrial DNA (mtDNA) rather than to nuclear DNA (*1–3*), mtDNA is considered to be their major cellular target. It has been hypothesized that the resultant somatic mutations in mtDNA play a causal role in oncogenic transformation (*3*). Many subsequent studies have supported the idea of preferential accumulation of somatic mutations in tumor mtDNAs (*4–9*) and their contribution to tumor growth (*10, 11*). However, the apparent high frequency of mtDNA mutations in tumors could be due either to their stochastic accumulation (*12, 13*) or to laboratory errors (*14*). Moreover, if mtDNA mutations induce oncogenic transformation, all the offspring of a mother carrying such mutations should

develop tumors due to the maternal inheritance of mtDNA (*15, 16*), but no bias toward maternal inheritance of tumor development has been reported. Nonetheless, it remains possible that mtDNA mutations are involved at a later stage of tumorigenesis, for example, in the development of metastatic potential. Recent studies demonstrated that dysfunction of the tricarboxylic acid cycle (TCA cycle) caused by mutations in nuclear DNA controls tumor phenotypes by the induction of a pseudo-hypoxic pathway under normoxic conditions (*17–19*). However, there has been no evidence of the involvement of mtDNA mutations in the development of metastatic potential or in the regulation of the pseudo-hypoxic pathway because of the difficulty of excluding possible involvement of nuclear DNA mutations in these processes (*20*).

We have examined the role of pathogenic mtDNA mutations in the development of tumor cell metastasis by studying two mouse tumor cell lines with different metastatic potentials (low metastatic P29 and high metastatic A11 cells) that originated from Lewis lung carcinoma (table S1) (*21–23*). We compared mitochondrial respiratory function by estimating the activities of respiratory complexes and found that P29 cells had normal activities, whereas A11 cells showed reduced ac-

tivity of complex I (NADH dehydrogenase) (Fig. 1A). Complex I defects were also observed in high metastatic fibrosarcoma B82M cells but not in high metastatic colon adenocarcinoma LuM1 cells (Fig. 1A), which suggests that metastatic tumors are not always associated with complex I defects.

Because complex I consists of subunits encoded by both nuclear DNA and mtDNA (*24*), it was necessary to determine which genome, nuclear or mitochondrial, was responsible for the complex I defects and whether the complex I defects were responsible for the high metastatic potential. We addressed these issues by complete reciprocal exchange of mtDNAs between P29 and A11 cells by means of cell fusion to isolate transmitochondrial cybrids (fig. S1A and table S2) and examined whether complex I defects and metastatic potentials were cotransferred with the mtDNA. The results showed that complex I activity decreased in the cybrids with A11 mtDNA, whereas those with P29 mtDNA showed normal activity, irrespective of whether their nuclear DNAs were derived from P29 or A11 cells (Fig. 1B). Thus, complex I defects in the cybrids with A11 mtDNA appear to result from pathogenic mutations in their mtDNA, not in their nuclear DNA. We then examined the metastatic potential of the cybrids by inoculating them into a tail vein (to test "experimental" metastasis) and under the skin (to test "spontaneous" metastasis) of C57BL/6 mice and counting the number of nodules formed in the lung. Cybrids with A11 mtDNA acquired high metastatic potential, whereas cybrids with P29 mtDNA lost metastatic potential (table S2). These observations suggest that complex I defects and high metastatic potential are transferred simultaneously with the transfer of mtDNA from the A11 cells, whereas normal complex I activity and low metastatic potential are transferred simultaneously with the transfer of mtDNA from P29 cells. The mtDNA of A11 cells is therefore likely to harbor a mutation(s) responsible for complex I defects and metastasis.

We next examined whether these findings could be generalized to additional tumor cell lines. In these experiments, we transferred mtDNA from A11 cells into fibrosarcoma B82 cells with low metastatic potential and normal complex I activity, resulting in isolation of B82mtA11 cybrids (table S2). Conversely, we transferred mtDNA from B82M cells, which are derived from B82 cells but express high metastatic potential and com-

[1]Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan. [2]Tsukuba Advanced Research Alliance, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan. [3]Japan Society for the Promotion of Science (JSPS), 8 Ichiban-cho, Chiyoda-ku, Tokyo 102-8472, Japan. [4]Division of Chemotherapy, Chiba Cancer Center Research Institute, 666-2 Nitona, Chuo-ku, Chiba 260-8717, Japan. [5]Shimane University Faculty of Medicine, 89-1 Enya-cho, Izumo, Shimane 693-8501, Japan.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: jih45@sakura.cc.tsukuba.ac.jp

plex I defects, into low metastatic P29 cells, resulting in isolation of P29mtB82M cybrids. Both B82MmtA11 and P29mtB82M cybrids acquired complex I defects (Fig. 1C) and high metastatic potential (table S2), which suggests the cotransfer of these phenotypes and the mtDNAs from high to low metastatic cells of different tumor types.

Notably, transfer of mtDNA from high metastatic A11 and B82M cells into nontransformed NIH3T3 cells did not induce tumorigenicity and metastatic potential in the resultant NIHmtA11 and NIHmtB82M cybrids (fig. S2A and table S2). Thus, pathogenic mtDNA mutations that induce complex I defects are present in A11 and

B82M cells and control development of metastases; however, these mutations do not control the development of tumorigenicity and metastasis, at least in nontransformed NIH3T3 cells.

To identify the pathogenic mtDNA mutations that induced complex I defects and high metastatic potential in A11 and B82M cells, we compared the whole mtDNA sequences between P29 and A11 cells and between B82 and B82M cells. We conclude that a missense G13997A mutation in the A11 cells and a frame-shift 13885insC mutation in the B82M cells, both within the *ND6* (NADH dehydrogenase subunit 6) gene, are the pathogenic mutations that induce complex I defects, because these are the only mutations exclusively observed in the mtDNA of the high metastatic A11 and B82M cells (Table 1). Restriction enzyme digestion of the polymerase chain reaction products amplified using mismatched primers suggests complete and reciprocal replacement of parental mtDNAs in our cybrids (fig. S3).

We next explored how the mutated mtDNA and resultant complex I defects regulate metastasis. Because complex I defects may lead to overproduction of reactive oxygen species (ROS) (24, 25), we estimated the amounts of ROS (fig. S4), and found that the cybrids with the mutated mtDNA from A11 cells showed enhanced ROS production, whereas the cybrids without the mutated mtDNA from P29 cells did not (fig. S4B). Such cotransfer of ROS-producing properties to the cybrids along with the transfer of mtDNA with or without the mutation suggests that ROS overproduction is due to the G13997A mutation. ROS overproduction was also observed in the P29mtB82M and B82MmtA11 cybrids (fig. S4C).

How does ROS overproduction regulate metastasis, and which nuclear genes (if any) are involved in this process? We have reported previously (22, 26) that A11 cells, but not P29 cells, show resistance to hypoxia-induced apoptosis, accompanied by up-regulation of antiapoptotic MCL-1 (myeloid cell leukemia–1). Moreover, A11 cells showed higher expression levels of two genes associated with neoangiogenesis, HIF-1α (hypoxia-inducible factor–1α) and VEGF (vascular endothelial growth factor), in comparison
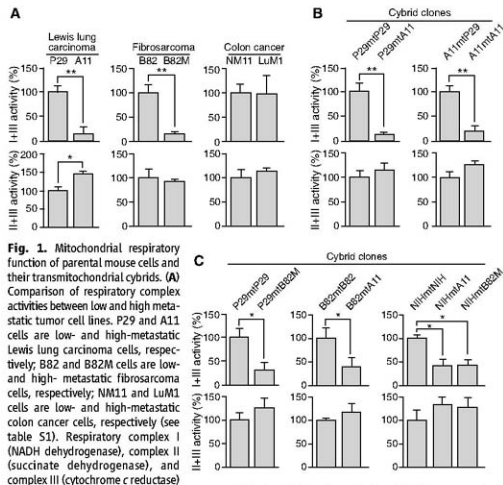
Fig. 1. Mitochondrial respiratory function of parental mouse cells and their transmitochondrial cybrids. (**A**) Comparison of respiratory complex activities between low and high metastatic tumor cell lines. P29 and A11 cells are low- and high-metastatic Lewis lung carcinoma cells, respectively; B82 and B82M cells are low- and high- metastatic fibrosarcoma cells, respectively; NM11 and LuM1 cells are low- and high-metastatic colon cancer cells, respectively (see table S1). Respiratory complex I (NADH dehydrogenase), complex II (succinate dehydrogenase), and complex III (cytochrome c reductase)

are components of the electron-transport chain located in the mitochondrial inner membrane. Mitochondrial respiratory function was examined by estimating their activities. Because the activity of complexes II+III is normal in the A11 and B82M cells, the reduced activity of complexes I+III exclusively observed in the A11 and B82M cells should represent complex I defects. (**B**) Comparison of respiratory complex activities of the cybrids with mtDNA exchanged between low-metastatic P29 and high-metastatic A11 Lewis lung carcinoma cells. (**C**) Comparison of respiratory complex activities of the cybrids with mtDNA transferred from different types of tumor cells expressing high metastatic potential. Bars represent the mean ± SD (n = 3). *P < 0.05; **P < 0.01.

Table 1. Identification of pathogenic mutations by comparison of mtDNA sequences between low- and high-metastatic mouse tumor cells.

Position	Gene	Amino acid change	Mouse strain C57BL/6	Cell lines					
				P29	A11	L929	B82	B82M	NIH3T3
T6589C	*COI*	V421A	T	T	T	C	C	C	T
G9348A	*COIII*	V248I	G	G	G	A	A	A	G
T9461C	*ND3*	Silent	T	C	C	C	C	C	G
9821-PolyA	*tRNA^Arg*	–	8A	9A	9A	10A	10A	10A	10A
C11493A	*ND4*	P443T	C	A	C	A	A	A	G
A13672T	*ND6*	Silent	A	T	T	A	A	A	A
13885insC	*ND6*	Frame-shift	–	–	–	–	–	C*	A
G13997A	*ND6*	P25L	G	G	A*	G	G	G	G
Accession No.			AY172335	EU312160	EU312161	AJ489607	EU315229	EU315228	AY999076

*The G13997A mutation in *ND6* is a missense mutation that changes the amino acid proline to leucine at a site that is highly conserved throughout vertebrates. The 13885insC mutation in *ND6* is a frame-shift mutation that has been previously reported as a pathogenic mutation that induces substantial complex I defects in some sublines of an L929 fibroblast cell line and A9 cells (23).

with P29 cells (27). Thus, we focused here on the expression of these three nuclear-coded genes. We found that up-regulation of the MCL-1, HIF-1α, and VEGF was cotransferred when mutant mtDNA was transferred from A11 cells to the P29mtA11 and A11mtA11 cybrids. Down-regulation of three genes was cotransferred when wild-type mtDNA was transferred from P29 cells to the P29mtP29 and A11mtP29 cybrids (Fig. 2).

Therefore, the mutated mtDNA and the resultant complex I defects induce up-regulation of the MCL-1, HIF-1α, and VEGF genes and are associated with high metastatic potential (fig. S1B). Gene expression profiling to compare P29mtP29 with P29mtA11 and A11mtP29 with A11mtA11 showed consistent up-regulation of other genes possibly related to metastasis in the cybrids with A11 mtDNA (table S3), which suggests involve-

ment of additional genes in the mtDNA-mediated effects on metastasis.

To obtain direct evidence that ROS over-production caused by the mutated mtDNA from A11 cells is responsible for high metastatic potential, we treated the P29mtA11 cybrids with ROS scavengers and examined their effects on the amounts of ROS and on the expression of the genes and the phenotypes related to metastasis. N-acetylcysteine (NAC), which has been used as an anticancer agent in preclinical models, was used as one ROS scavenger. The results showed that treatment of the cybrids with NAC in cell culture reduced the amount of ROS (Fig. 3A) and down-regulated MCL-1 (Fig. 3B). Moreover, pretreatment of the cybrids with NAC reduced their metastatic potential in two mouse models (Fig. 3C). Similar results were obtained by treatment with another ROS scavenger, Ebselen, which is a mimic of glutathione peroxidase (Fig. 3). Thus, ROS overproduction caused by the mutated mtDNA induces a high metastatic potential, at least in part, by up-regulation of MCL-1. This idea is supported by the finding that down-regulation of MCL-1 in P29mtA11 cybrids by small interfering RNA also suppressed their metastatic potential (fig. S5). Moreover, NAC treatment suppressed the metastatic potential without reducing glycolytic activity (fig. S6), which suggests that metastasis is not caused by up-regulation of glycolysis.

Contribution of mtDNA to tumor cell metastasis can be extended to human tumors, because the transfer of mtDNA from human breast cancer MDA-MB-231 cells expressing high metastatic potential to low metastatic HeLa cells induces
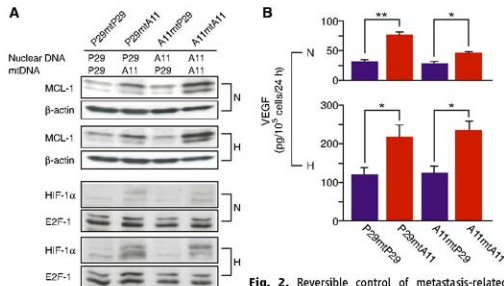
**Fig. 2.** Reversible control of metastasis-related nuclear gene expression by mtDNA. (**A**) Expressions of nuclear-coded MCL-1 and HIF-1α and (**B**) VEGF under normoxia (N) and hypoxia (H). As loading controls in the Western blots, we used β-actin for MCL-1 and E2F-1 for HIF-1α (A). In (B), blue bars represent cybrids carrying mtDNA from P29 cells (P29mtP29 and A11mtP29), and red bars represent cybrids carrying mtDNA from A11 cells (A11mtA11 and P29mtA11). Bars represent the mean ± SD (n = 3). *P < 0.01; **P < 0.001.
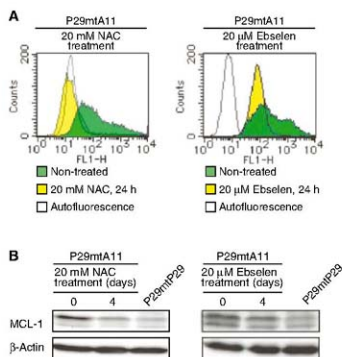
Fig. 3. Suppression of metastasis by treatment of the P29mtA11 cybrids with ROS scavengers. (**A**) Effects of NAC and Ebselen treatments on the amounts of ROS. The P29mtA11 cybrids (1 × 10^6 cells) treated with 5 μM dichlorofluorescein diacetate were subjected to fluorescence-activated cell sorting (FACS) analysis for quantitative estimation of ROS (H_2O_2). FACS was carried out before (green) and after (yellow) 24 hours of treatment of the cybrids with 20 mM NAC or 20 μM Ebselen. (**B**) Effects of NAC and Ebselen treatments on MCL-1 expression. Western blot analysis of MCL-1 was carried out before and after the treatment of P29mtA11 cybrids with 20 mM NAC or 20 μM Ebselen for 4 days. β-actin served as the loading control. (**C**) Effects of NAC and Ebselen treatments on metastatic potential. The P29mtA11 cybrids pretreated for 4 days with 20 mM NAC or with 20 μM Ebselen were injected into the tail vein of C57BL/6 mice to test the experimental metastatic potential. To examine the effect of NAC administration on spontaneous metastatic potential, P29mtA11 cybrids without NAC pretreatment were injected subcutaneously into C57BL/6 mice, which subsequently received 10 mg/ml NAC in drinking water ad libitum. Bars represent the mean ± SD (n = 6). *P < 0.05; **P < 0.01.
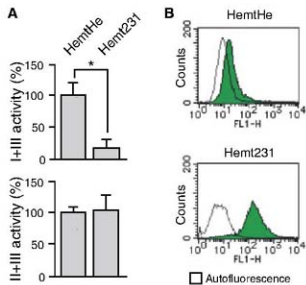
Fig. 4. Induction of high metastatic potential in HeLa cells by introduction of mtDNA from human breast adenocarcinoma MDA-MB-231 cells expressing high metastatic potential. Induction of **(A)** complex I defects, **(B)** ROS overproduction, and **(C)** high metastatic potential in the Hemt231 cybrids. HemtHe, HemtHe cybrids carrying nuclear DNA from ρ^0 HeLa cells and mtDNA from wild-type HeLa cells; Hemt231, Hemt231 cybrids carrying nuclear DNA from ρ^0 HeLa cells and mtDNA from MDA-MB-231 cells. Bars represent the mean ± SD ($n = 5$). *$P < 0.05$.

complex I defects, increased ROS production, and high metastatic potential in HeLa cells (Fig. 4). These observations suggest that the mtDNA in MDA-MB-231 cells can promote metastasis, although we have not done the mtDNA sequencing. Therefore, the metastatic potential of all the mouse and human tumor cell lines that we examined was greatly enhanced by exchanging their endogenous mtDNA with mutant mtDNA that induces complex I–mediated ROS overproduction. Recent reports showed that a pathogenic mutation in the *ATP6* gene of human mtDNA generated ROS and enhanced tumor growth (10, 11). However, in our experiments, the enhanced growth rate of primary tumors did not necessarily correlate with expression of the high metastatic potential in mouse tumors (fig. S2B).

This study partially resolves the debate on the relevance of mtDNA mutations in tumors (4–14) by showing that mutations in mtDNA can control the metastatic potential of certain tumor cells but that they do not confer tumorigenic potential to nontransformed mouse NIH3T3 cells. Moreover, reversible regulation of metastasis by the exchange of mtDNA between P29 and A11 cells and by treatment with ROS scavengers suggests that metastasis of these cells is regulated by ROS-mediated reversible up-regulation of nuclear genes but not by ROS-mediated acceleration of genetic instability. The mtDNA-mediated reversible control of metastasis, therefore, reveals a novel function of mtDNA and suggests that in such cases ROS scavengers may be therapeutically effective in suppressing metastasis.

References and Notes

1. J. A. Allen, M. M. Coombs, *Nature* **287**, 244 (1980).
2. J. M. Backer, I. B. Weinstein, *Science* **209**, 297 (1980).
3. J. W. Shay, H. Werbin, *Mutat. Res.* **186**, 149 (1987).
4. K. Polyak *et al.*, *Nat. Genet.* **20**, 291 (1998).
5. M. S. Fliss *et al.*, *Science* **287**, 2017 (2000).
6. J. S. Penta, F. M. Johnson, J. T. Wachsman, W. C. Copeland, *Mutat. Res.* **488**, 119 (2001).
7. R. W. Taylor, D. M. Turnbull, *Nat. Rev. Genet.* **6**, 389 (2005).
8. A. A. M. Czarnecka, P. Golik, E. Bartnik, *J. Appl. Genet.* **47**, 67 (2006).
9. M. E. Gallardo *et al.*, *Hum. Mutat.* **27**, 575 (2006).
10. Y. Shidara *et al.*, *Cancer Res.* **65**, 1655 (2005).
11. J. A. Petros *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 719 (2005).
12. H. A. Coller *et al.*, *Nat. Genet.* **28**, 147 (2001).
13. R. W. Taylor *et al.*, *J. Clin. Invest.* **112**, 1351 (2003).
14. A. Salas *et al.*, *PLoS Med.* **2**, e296 (2005).
15. H. Kaneda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4542 (1995).
16. H. Shitara, J.-I. Hayashi, S. Takahama, H. Kaneda, H. Yonekawa, *Genetics* **148**, 851 (1998).
17. B. E. Baysal *et al.*, *Science* **287**, 848 (2000).
18. S. Niemann, U. Muller, *Nat. Genet.* **26**, 268 (2000).
19. E. Gottlieb, I. P. Tomlinson, *Nat. Rev. Cancer* **5**, 857 (2005).
20. L. H. Augenlicht, B. Heerdt, *Nat. Genet.* **28**, 104 (2001).
21. K. Takenaga, Y. Nakamura, S. Sakiyama, *Oncogene* **14**, 331 (1997).
22. M. Takasu, Y. Tada, J. O. Wang, M. Tagawa, K. Takenaga, *Clin. Exp. Metastasis* **17**, 409 (1999).
23. Materials and methods are available as supporting material on *Science* Online.
24. D. C. Wallace, *Science* **283**, 1482 (1999).
25. S. De Flora *et al.*, *Int. J. Cancer* **67**, 842 (1996).
26. N. Koshikawa, C. Maejima, K. Miyazaki, A. Nakagawara, K. Takenaga, *Oncogene* **25**, 917 (2006).
27. N. Koshikawa, A. Iyozumi, M. Gassmann, K. Takenaga, *Oncogene* **22**, 6717 (2003).
28. This work was supported by Grants-in-Aid for Scientific Research (S) from the Japan Society for Promotion of Science (JSPS), by Grants-in-Aid for Creative Scientific Research from JSPS, and by Grants-in-Aid for Scientific Research on Priority Areas from The Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) to J.-I.H. This work was also supported by grants for a Research Fellowship from JSPS for Young Scientists to K.I. and by Grants-in-Aid for Third Term Comprehensive Control Research for Cancer from the Ministry of Health, Labour, and Welfare and for Scientific Research from MEXT to K.T. All animal experiments were performed in compliance with the institutional guidelines (Chiba Cancer Center and University of Tsukuba) for the care and use of laboratory animals.

In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish

Scott T. Laughlin,[1]* Jeremy M. Baskin,[1]* Sharon L. Amacher,[2] Carolyn R. Bertozzi[1,2,3,4]†

Glycans are attractive targets for molecular imaging but have been inaccessible because of their incompatibility with genetically encoded reporters. We demonstrated the noninvasive imaging of glycans in live developing zebrafish, using a chemical reporter strategy. Zebrafish embryos were treated with an unnatural sugar to metabolically label their cell-surface glycans with azides. Subsequently, the embryos were reacted with fluorophore conjugates by means of copper-free click chemistry, enabling the visualization of glycans in vivo at subcellular resolution during development. At 60 hours after fertilization, we observed an increase in de novo glycan biosynthesis in the jaw region, pectoral fins, and olfactory organs. Using a multicolor detection strategy, we performed a spatiotemporal analysis of glycan expression and trafficking and identified patterns that would be undetectable with conventional molecular imaging approaches.

The cell-surface glycome is a rich source of information that reports on the cell's physiological state. For example, changes in glycan structures serve as markers of altered gene expression during development (1) and disease progression (2). The dynamics of glycans at the plasma membrane reflect the activity of the cell's secretory machinery (3), and their relative abundances report on flux in metabolic pathways inside the cell (4). Glycans are therefore attractive targets for in vivo imaging but have been inaccessible because of their incompatibility with genetically encoded reporters (5).

To image glycans in vivo, we employed a strategy in which an azide is introduced into target biomolecules, priming them for selective covalent reaction with fluorescent probes (5). The azide is small, stable in biological systems, and selectively reactive with phosphines or activated alkynes. Previously, the Staudinger ligation (6, 7) or copper-catalyzed click chemistry (8, 9) have been used to detect azide-labeled biomolecules on cells ex vivo. However, in vivo

imaging of dynamic biological processes using these chemistries could be complicated by slow reaction kinetics or reagent toxicity. The copper-free click reaction of azides with difluorinated cyclooctyne (DIFO) reagents (10) overcomes these limitations, suggesting its potential application to in vivo imaging.

We chose zebrafish as a model organism because of its well-defined developmental program (11), emerging disease models (12), and amenability to optical imaging. The metabolic substrate peracetylated N-azidoacetylgalactosamine (Ac₄GalNAz) was selected on the basis of its known incorporation into mucin-type O-linked glycoproteins in mammalian cells and mice via the N-acetylgalactosamine (GalNAc) salvage pathway (13, 14) (fig. S1). We envisioned an imaging experiment (Fig. 1A) in which zebrafish embryos are incubated with Ac₄GalNAz and their glycans are visualized by reaction with DIFO-fluorophore conjugates (fig. S2).

Before performing imaging experiments, we confirmed that the zebrafish glycan biosynthetic enzymes are permissive of the unnatural sugar. The zebrafish cell line ZF4 (15) was incubated with various doses of Ac₄GalNAz, reacted with a DIFO–Alexa Fluor 488 conjugate (DIFO-488, fig. S2), and analyzed by flow cytometry (Fig. 1B). Robust dose-dependent metabolic labeling was observed, similar to that of mammalian cells (13, 14). We further characterized the azide-labeled cell lysates by treatment with a DIFO–Flag conjugate (10). The observed high-molecular-weight species were consistent with labeled glycoproteins (fig. S3). We then purified the Flag-containing species (16) and identified several glycoproteins (β-hexosaminidase, β-integrin 1b, lysosome-associated membrane protein, nicastrin, scavenger receptor B, and Thy1) with known (17–19) or predicted (20) sites of mucin-type O-linked glycosylation (fig. S4). We concluded that Ac₄GalNAz was metabolically incorporated into glycoproteins in zebrafish-derived cells.

We next evaluated Ac₄GalNAz labeling in vivo. Zebrafish embryos were incubated in media containing either Ac₄GalNAz or, as a control, peracetylated GalNAc (Ac₄GalNAc) from 3 to 120 hours post-fertilization (hpf). Whole-animal lysates were then reacted with a phosphine-Flag probe (21) (fig. S2) and analyzed (Fig. 1C). The labeled glycoproteins were refractory to digestion with peptide N-glycosidase F or chondroitinase ABC (fig. S5), which suggests that GalNAz is primarily incorporated into mucin-type O-linked glycoproteins.

To image azide-labeled glycans in vivo, we incubated zebrafish embryos with either Ac₄GalNAz or Ac₄GalNAc from 3 to 72 hpf and then reacted the embryos with a DIFO–Alexa Fluor 647 conjugate (DIFO-647, fig. S2). Robust fluorescence was observed with virtually no background (Fig. 2A). Even after a 1-min reaction with DIFO-647, the Ac₄GalNAz-treated embryos displayed substantial fluorescence that increased in a time-dependent manner (Fig. 2B). We observed no toxicity or developmental abnormalities resulting from treatment with Ac₄GalNAz or any DIFO reagents (fig. S6 and supporting online material text).

We then assessed global patterns of glycosylation by incubating embryos with Ac₄GalNAz starting at 3 hpf, followed by reaction with DIFO-647 at 12-hour intervals over a 5-day period. We observed azide-labeled glycans as early as 24 hpf (Fig. 2C and fig. S6). Starting at 60 hpf and continuing until at least 72 hpf, we observed a burst in fluorescence intensity in the jaw region, pectoral fins, and olfactory organs (Fig. 2, D and E). Thus, we focused on 60 to 72 hpf for more


Fig. 1. Ac₄GalNAz is metabolically incorporated into zebrafish glycans. (**A**) Schematic depicting the use of metabolic labeling with Ac₄GalNAz and copper-free click chemistry using DIFO probes for noninvasive imaging of glycans during zebrafish development. (**B**) Flow cytometry analysis of ZF4 cells metabolically labeled with Ac₄GalNAz. ZF4 cells were incubated with Ac₄GalNAz (0 to 100 μM, 3 days) and subsequently reacted with DIFO-488 (10 μM, 1 hour). Error bars represent the standard deviation from three replicate samples. (**C**) Immunoblot analysis of lysates from zebrafish embryos at 120 hpf incubated with Ac₄GalNAc (Ac) or Ac₄GalNAz (Az), probed with horseradish peroxidase–conjugated antibody to Flag (top panel) or antibody to β-tubulin (bottom panel).
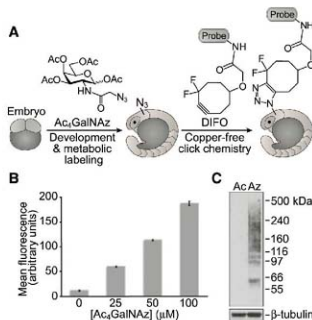
Fig. 2. In vivo imaging of glycans during zebrafish development. (**A** and **B**) Zebrafish embryos were metabolically labeled with Ac₄GalNAz (Az) or Ac₄GalNAc (Ac) starting at 3 hpf. (A) Embryos were reacted at 72 hpf with DIFO-647 for 1 hour. The right panel indicates an exposure time that is 20 times longer than that in the other two panels. (B) Embryos were reacted at 72 hpf with DIFO-647 for 1 to 60 min. Asterisks denote autofluorescence. (**C**) Zebrafish embryos incubated with Ac₄GalNAz or Ac₄GalNAc (fig. S6) starting at 3 hpf were reacted with DIFO-647 at 24 hpf and subsequently at 12-hour intervals, viewed laterally and ventrally (alternating panels). (**D** and **E**) Zebrafish from (C) imaged at higher magnification at 60 hpf (D) or 72 hpf (E), viewed laterally (left panels) and ventrally (right panels). Solid arrowhead, olfactory organ; open arrowhead, pectoral fin. Dotted line indicates the pharyngeal epidermis in the jaw region. Scale bars in (A) and (C), 500 μm; in (B), (D), and (E), 200 μm.
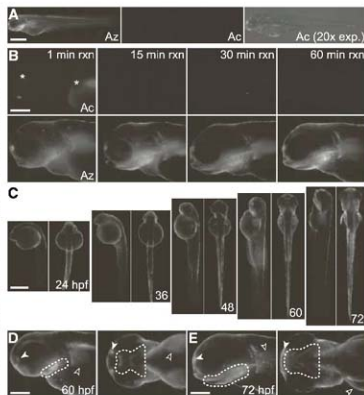
¹Department of Chemistry, University of California, Berkeley, CA 94720, USA. ²Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. ³Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA. ⁴The Molecular Foundry, Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: crb@berkeley.edu

Fig. 3. Identification of temporally distinct glycan populations during zebrafish development using two-color labeling. Zebrafish embryos metabolically labeled with Ac₄GalNAz from 3 to 60 hpf were reacted with DIFO-647 between 60 and 61 hpf and then reacted with DIFO-488 either between 61 and 62 hpf [(A) to (D)] or, after an additional 1 hour of metabolic labeling with Ac₄GalNAz, between 62 and 63 hpf [(E) to (I)]. Control embryos incubated with Ac₄GalNAc and otherwise reacted with the same DIFO-fluorophore probes are shown in figs. S9 and S10. (**A**) Brightfield image of a frontal view. (**B**) z-projection (left panel) and x-projection (right panel) fluorescence images of the mouth region. (**C**) Brightfield image of a lateral view. (**D**) Single z-plane fluorescence image of the pectoral fin region. (**E**) Brightfield image of a ventral view of an embryo at 63 hpf. (**F**) Single z-plane fluorescence image of (E) displaying intense DIFO-488 fluorescence but not DIFO-647 fluorescence. (**G**) Left panel, single z-plane fluorescence image of the jaw region; middle and right panels, z-projection (middle panel) and x-projection (right panel) fluorescence images of the region highlighted in the left panel. (**H**) z-projection (left panel) and y-projection (right panel) fluorescence images of the mouth. (**I**) z-projection fluorescence image of the olfactory organ. Highlighted are the olfactory epithelium (oe) and olfactory pit (op) regions. In (B), (D), and (F) to (I), red is DIFO-647 (60 to 61 hpf) and green is DIFO-488 [61 to 62 hpf in (B) and (D) and 62 to 63 hpf in (F) to (I)]. Scale bars in (A), (C), (E), and (F), 100 μm; in (B), (D), (G) (left panel), (H), and (I), 10 μm; in (G) (middle and right panels), 5 μm.
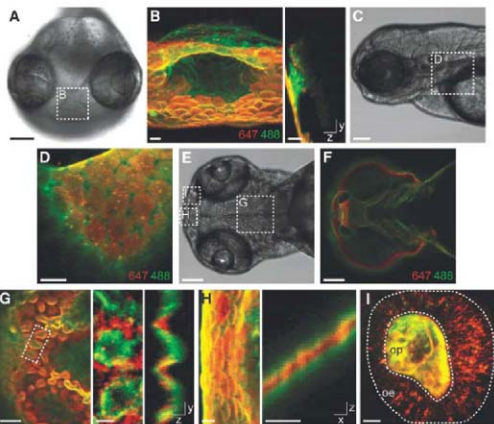
Fig. 4. Spatiotemporal analysis of de novo glycan biosynthesis during zebrafish development between 60 and 72 hpf. Zebrafish embryos metabolically labeled with Ac₄GalNAz from 3 to 60 hpf were reacted with DIFO-647 between 60 and 61 hpf, metabolically labeled with Ac₄GalNAz for 1 hour, and reacted with DIFO-488 between 62 and 63 hpf. The embryos were metabolically labeled with Ac₄GalNAz for an additional 9 hours and then reacted with DIFO-555 between 72 and 73 hpf. (**A**) z-projection fluorescence image of a lateral view. (**B**) Single z-plane fluorescence images of the region highlighted in (A). (**C**) Single z-plane fluorescence image of a ventral view of the jaw region. (**D**) Left panel, z-projection fluorescence image of cells in the region highlighted in (C); middle and right panels, z-projection (middle panel) and x-projection (right panel) fluorescence images of the cells highlighted in the left panel (white dashed rectangle). (**E**) z-projection (left panel) and x-projection (right panel) fluorescence images of kinocilia. (**F**) z-projection fluorescence image of the olfactory organ. Highlighted are the olfactory epithelium (oe) and olfactory pit (op) regions. Blue, DIFO-647 (60 to 61 hpf); green, DIFO-488 (62 to 63 hpf); red, DIFO-555 (72 to 73 hpf). Scale bars in (A) and (C), 100 μm; in (B), 25 μm; in (D) and (F), 10 μm; in (E), 5 μm.
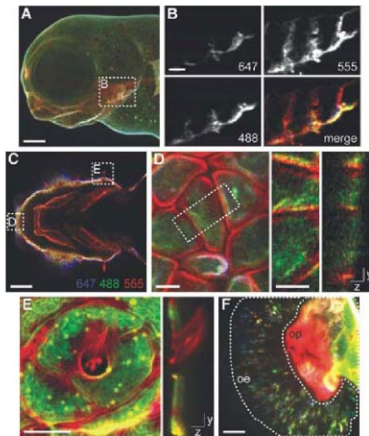
detailed studies of glycan expression and dynamics in these structures.

We sought to resolve temporally distinct populations of glycans using two- and three-color detection experiments (fig. S7). Embryos labeled with Ac₄GalNAz were reacted with DIFO-647 at 60 hpf to visualize the cell-surface glycans exposed at that time point. Because the fluorophore cannot penetrate cells (10), nascent azide-labeled glycans trafficking through the secretory pathway remained unreacted. In order to distinguish these newly synthesized glycans from the previously reacted population, we treated the embryos with tris-(2-carboxyethyl)phosphine (TCEP) to quench unreacted cell-surface azides and then reacted the embryos with a second fluorophore, DIFO-488 (fig. S8). After the procedure, the "old" glycans could be visualized by DIFO-647 fluorescence and the "new" glycans by DIFO-488 fluorescence (fig. S7).

Throughout the organism, we observed zones of de novo glycan biosynthesis (Fig. 3, A to D, and fig. S9). For example, the invagination of the mouth was labeled minimally by the first reaction but prominently by the second (Fig. 3, A and B, and movie S1), suggesting that although this structure was present at 60 hpf, its cells had only recently synthesized large amounts of GalNAz-labeled glycans. Further, we could readily distinguish plasma membrane–associated glycans from those that had been internalized by the cells. We noticed differential rates of endocytosis among cells throughout the embryo. In the eye and dorsal epithelium regions, prominent cell-surface

fluorescence was apparent from both DIFO reagents, suggesting a slow rate of glycan internalization (movie S2). However, in the pectoral fin, the old glycans detected with DIFO-647 (at 60 hpf) had been almost entirely internalized by the time the embryos were imaged, whereas the new glycans detected with DIFO-488 (at 62 hpf) were predominantly cell-surface–bound (Fig. 3, C and D).

To capture a broader spectrum of newly synthesized glycans, we expanded the period between the two DIFO-fluorophore reactions to 2 hours. Using this protocol, we observed intense labeling of the pharyngeal epidermis in the jaw region that was derived from the second reaction (DIFO-488) but not from the first (DIFO-647) (Fig. 3, E and F; fig. S10; and movie S3). In caudal regions of the pharyngeal epidermis that were labeled during both reactions, we noticed a corrugated distribution of glycans in which old glycans were restricted to peaks at the extreme ventral surface and new glycans were produced in troughs projecting dorsally (Fig. 3G and movie S4). This corrugated pattern was not observed in other regions of the animal (for example, Fig. 3H and fig. S11). Analysis of the olfactory organ also revealed a clear spatial distinction between the old and new glycans. The more recently produced glycans were predominantly localized in the olfactory pit, whereas older glycans were present in both the olfactory pit and epithelium (Fig. 3I and movie S5). The order of treatment with the two DIFO-fluorophores had no effect on the observed patterns (fig. S12).

Finally, we expanded our analysis to encompass the period from 60 to 72 hpf using three DIFO-fluorophore conjugates (DIFO-647, DIFO-488, and DIFO-555; fig. S2). A population of doubly reacted embryos was generated as before but then quenched with TCEP a second time, allowed to develop for 9 hours in Ac₄GalNAz, and finally labeled with DIFO-555 (fig. S7). Glycan production between 63 and 72 hpf was evident throughout the jaw region (Fig. 4, A to C, and movies S6 and S7), which was labeled robustly with DIFO-555 but minimally with DIFO-647 and –488. In contrast, cells analyzed from the extreme rostral region displayed DIFO-555 fluorescence on the cell membrane as well as intracellular DIFO-647 and –488 fluorescence derived from internalized older glycans (Fig. 4D). Additionally, the kinocilia of mechanosensory hair cells surrounding the head of the embryo displayed fluorescence from DIFO-555 but not from DIFO-647 or –488 (Fig. 4E and movie S8). In contrast, adjacent epithelial cells displayed fluorescence from all three DIFO reagents, indicating their maturation during an earlier period in development. We also observed newer, DIFO-555–labeled glycans on cilia in the olfactory pit, whereas the majority of the DIFO-647 and –488 fluorescence was localized in the olfactory epithelium (Fig. 4F, fig. S13, and movie S9). Thus, olfactory pit glycans may be rapidly degraded or released from the embryo; alternatively, glycans produced in the olfactory pit may migrate to the olfactory epithelium.

Metabolic labeling with Ac₄GalNAz followed by detection via copper-free click chemistry re-

vealed differences in the cell-surface expression, intracellular trafficking, and tissue distribution of glycans throughout zebrafish embryogenesis. This approach may be generalized to alternative imaging modalities and to other biomolecules (5) [for example, sialic acids can be imaged with N-azidoacetylmannosamine (fig. S14)].

References and Notes
1. R. S. Haltiwanger, J. B. Lowe, *Annu. Rev. Biochem.* **73**, 491 (2004).
2. K. Ohtsubo, J. D. Marth, *Cell* **126**, 855 (2006).
3. D. N. Hebert, S. C. Garman, M. Molinari, *Trends Cell Biol.* **15**, 364 (2005).
4. S. Wopereis, D. J. Lefeber, E. Morava, R. A. Wevers, *Clin. Chem.* **52**, 574 (2006).
5. J. A. Prescher, C. R. Bertozzi, *Nat. Chem. Biol.* **1**, 13 (2005).
6. J. A. Prescher, D. H. Dube, C. R. Bertozzi, *Nature* **430**, 873 (2004).
7. P. V. Chang, J. A. Prescher, M. J. Hangauer, C. R. Bertozzi, *J. Am. Chem. Soc.* **129**, 8400 (2007).
8. K. E. Beatty *et al.*, *Angew. Chem. Int. Ed.* **45**, 7364 (2006).
9. M. Sawa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12371 (2006).
10. J. M. Baskin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16793 (2007).
11. C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, *Dev. Dyn.* **203**, 253 (1995).
12. G. J. Lieschke, P. D. Currie, *Nat. Rev. Genet.* **8**, 353 (2007).
13. H. C. Hang, C. Yu, D. L. Kato, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14846 (2003).
14. D. H. Dube, J. A. Prescher, C. N. Quang, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4819 (2006).
15. W. Driever, Z. Rangini, *In Vitro Cell. Dev. Biol. Anim.* **29**, 749 (1993).
16. S. T. Laughlin *et al.*, *Methods Enzymol.* **415**, 230 (2006).
17. S. R. Carlson, P. O. Lycksell, M. Fukuda, *Arch. Biochem. Biophys.* **304**, 65 (1993).
18. U. Yabe, C. Sato, T. Matsuda, K. Kitajima, *J. Biol. Chem.* **278**, 13875 (2003).
19. M. Clement, J. Rocher, G. Loirand, J. Le Pendu, *J. Cell Sci.* **117**, 5059 (2004).
20. J. E. Hansen *et al.*, *Glycoconj. J.* **15**, 115 (1998).
21. K. L. Kilck, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 19 (2002).
22. We thank K. Blum, J. Codelli, E. Janus, and J. St. Hilaire for technical assistance and N. Agard, M. Boyce, P. Chang, J. Ngai, D. Raible, T. Schilling, and J. Seeliger for helpful discussions. This work was funded by grants to C.R.B. (GM058867) and S.L.A. (GM061952) from the NIH. J.M.B. was supported by NSF and National Defense Science and Engineering predoctoral fellowships.

Phosphorylation by p38 MAPK as an Alternative Pathway for GSK3β Inactivation

Tina M. Thornton,[1] Gustavo Pedraza-Alva,[1]* Bin Deng,[2,3] C. David Wood,[1] Alexander Aronshtam,[1] James L. Clements,[4] Guadalupe Sabio,[5] Roger J. Davis,[5] Dwight E. Matthews,[1,3] Bradley Doble,[6] Mercedes Rincon[1]†

Glycogen synthase kinase 3β (GSK3β) is involved in metabolism, neurodegeneration, and cancer. Inhibition of GSK3β activity is the primary mechanism that regulates this widely expressed active kinase. Although the protein kinase Akt inhibits GSK3β by phosphorylation at the N terminus, preventing Akt-mediated phosphorylation does not affect the cell-survival pathway activated through the GSK3β substrate β-catenin. Here, we show that p38 mitogen-activated protein kinase (MAPK) also inactivates GSK3β by direct phosphorylation at its C terminus, and this inactivation can lead to an accumulation of β-catenin. p38 MAPK–mediated phosphorylation of GSK3β occurs primarily in the brain and thymocytes. Activation of β-catenin–mediated signaling through GSK3β inhibition provides a potential mechanism for p38 MAPK–mediated survival in specific tissues.

The p38 mitogen-activated protein kinase (MAPK) is activated through phosphorylation primarily by MAPK kinase 3 (MKK3) and MKK6 in response to cellular stress and cytokines. The p38 MAPK pathway functions in the control of differentiation, the blockade of proliferation, and in the induction of apoptosis (1). It is also activated in response to DNA double-stranded breaks (DSBs) induced by ionizing irradiation or chemotherapeutic drugs, and it participates in the induction of a G₂/M cell-cycle checkpoint (2, 3). p38 MAPK can also promote survival (4–6) by unknown mechanisms. During T cell receptor β (TCRβ) rearrangement, V(D)J recombination-mediated DSBs also activate

p38 MAPK in immature thymocytes at the double negative 3 (DN3) stage of development (7, 8). The expression of a constitutively active mutant of MKK6 [MKK6(Glu)] in thymocytes of transgenic mice (MKK6 transgenic mice) activates a p53-mediated G₂/M phase cell-cycle checkpoint (8). Like recombination-activating gene (Rag) deficiency, persistent activation of p38 MAPK interferes with the differentiation of thymocytes beyond the DN3 stage. However, MKK6 transgenic thymocytes (but not Rag⁻/⁻ thymocytes) survive and accumulate in vivo (8), suggesting that p38 MAPK may also provide a survival signal. A gene expression profile analysis comparing Rag⁻/⁻ and MKK6 DN3 thymo-
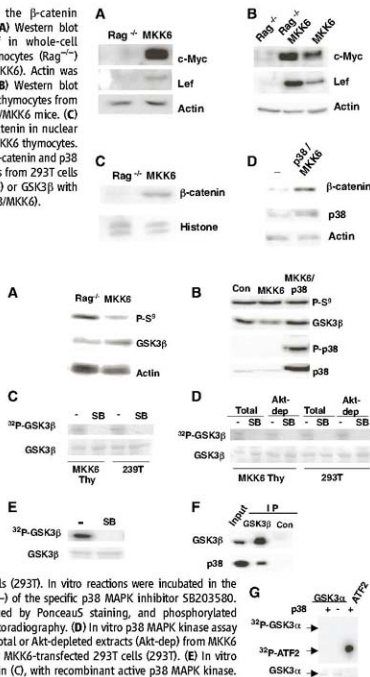
cytes revealed that the MKK6 DN3 thymocytes expressed more c-myc and lef (fig. S1) [two transcription factors associated with cell survival (9–11)] than did the Rag$^{-/-}$ thymocytes. The increased abundance of c-Myc and Lef proteins in the MKK6 transgenic thymocytes, compared with Rag$^{-/-}$ thymocytes, was confirmed by Western blot analysis (Fig. 1A) (12). Thymocytes from Rag$^{-/-}$ mice crossed with MKK6 transgenic (Rag$^{-/-}$ MKK6) mice contained higher amounts of c-Myc and Lef proteins than did Rag$^{-/-}$ thymocytes, indicating that the activation of p38 MAPK, but not the pre-TCR signals, contributes to the enhanced expression of these transcription factors (Fig. 1B). The c-myc and lef genes are targets of the β-catenin signaling pathway in certain contexts (13, 14). Nuclear accumulation of β-catenin was detected in MKK6 thymocytes, but not in Rag$^{-/-}$ thymocytes (Fig. 1C). Expression of constitutively active MKK6 in 293T cells was also sufficient to increase the amount of β-catenin protein (Fig. 1D), but this had no effect on β-catenin mRNA (fig. S2).

Phosphorylation of β-catenin by glycogen synthase kinase 3β (GSK3β) targets β-catenin for ubiquitination and subsequent degradation (15, 16). The best-characterized mechanism for the inactivation of GSK3β is through phosphorylation of its N terminus at Ser^9 by Akt (17). No increase was observed in the amount of phospho-Ser^9 GSK3β in MKK6 thymocytes compared with that in Rag$^{-/-}$ thymocytes (Fig. 2A). Similarly, no increase in phospho-Ser^9 was observed in 293T cells transfected with constitutively active MKK6 (Fig. 2B). Phosphorylation of Ser^9 was impaired by Wortmanin, an inhibitor of the PI3K-Akt pathway, but it was not affected by the pharmacological inhibitor of p38 MAPK SB203580 (fig. S3). Thus, p38 MAPK does not appear to regulate the Akt-mediated phosphorylation of GSK3β on Ser^9. p38 MAPK immunoprecipitated from MKK6 thymocytes or MKK6-transfected 293T cells phosphorylated recombinant catalytically inactive GSK3β in vitro, and this phosphorylation was blocked by the p38 MAPK inhibitor (Fig. 2C). No Akt was detected in p38 MAPK immunoprecipitates, and no p38 MAPK was detected in Akt immunoprecipitates (fig. S4), ruling out the presence of residual AKT associated with p38 MAPK. The depletion of Akt before immunoprecipitating p38 MAPK did

Fig. 1. Regulation of the β-catenin pathway by p38 MAPK. **(A)** Western blot showing c-Myc and Lef in whole-cell extracts from Rag$^{-/-}$ thymocytes (Rag$^{-/-}$) and MKK6 thymocytes (MKK6). Actin was examined as a control. **(B)** Western blot showing c-Myc and Lef in thymocytes from Rag$^{-/-}$, MKK6, and Rag$^{-/-}$/MKK6 mice. **(C)** Western blot showing β-catenin in nuclear extracts from Rag$^{-/-}$ and MKK6 thymocytes. **(D)** Western blot showing β-catenin and p38 MAPK in whole-cell extracts from 293T cells transfected with GSK3β (−) or GSK3β with p38 MAPK and MKK6 (p38/MKK6).

Fig. 2. Direct phosphorylation of GSK3β by p38 MAPK. **(A)** Western blot showing absence of phospho-Ser^9 GSK3β (P-S^9) and total GSK3β in Rag$^{-/-}$ and MKK6 thymocytes. **(B)** Western blot showing P-Ser^9 GSK3β, GSK3β, phospho-p38 MAPK (P-p38), and p38 MAPK in 293T cells transfected with an empty vector (Con), MKK6 alone, or MKK6 and p38 MAPK (MKK6/p38). **(C)** In vitro p38 MAPK assay with inactive recombinant GSK3β as the substrate and p38 MAPK immunoprecipitated from MKK6-thymocytes (MKK6 Thy) or

MKK6-transfected 293T cells (293T). In vitro reactions were incubated in the presence (SB) or absence (−) of the specific p38 MAPK inhibitor SB203580. Total GSK3β was visualized by Ponceau S staining, and phosphorylated GSK3β was detected by autoradiography. **(D)** In vitro p38 MAPK kinase assay as described in (C), using total or Akt-depleted extracts (Akt-dep) from MKK6 thymocytes (MKK6 Thy) or MKK6-transfected 293T cells (293T). **(E)** In vitro kinase assay as described in (C), with recombinant active p38 MAPK kinase. **(F)** Western blot showing GSK3β and p38 MAPK in the GSK3β and p21 (Con) immunoprecipitates (IP) and whole-cell extracts from MKK6 thymocytes (Input). **(G)** In vitro kinase assay for recombinant active p38 MAPK kinase using catalytically inactive GSK3α as a substrate. Phosphorylation of activating transcription factor 2 (ATF2) was examined as a positive control.

not affect the phosphorylation of GSK3β (Fig. 2D). A purified recombinant activated p38 MAPK also phosphorylated GSK3β in vitro, and this phosphorylation was blocked by SB203580 (Fig. 2E). Coimmunoprecipitation analysis showed that p38 MAPK was present in GSK3β immunoprecipitates from MKK6 thymocytes (Fig. 2F) and 293T cells (fig. S5). Thus, p38 MAPK physically associates with and phosphorylates GSK3β at a Ser^9-independent residue. Although GSK3α and GSK3β are thought to be similarly regulated and can compensate for each other for some functions (18, 19), GSK3α was not phosphorylated by recombinant p38 MAPK in vitro (Fig. 2G).

The MAPK extracellular signal–regulated protein kinase (ERK) phosphorylates Thr^{43} of GSK3β (20) but does not affect GSK3β activity. Although SerPro or ThrPro motifs recognized by ERK are also recognized by other MAPK groups, p38 MAPK was still able to partially phosphorylate a GSK3β-T^{43}A mutant (Fig. 3A), suggesting the existence of additional phosphorylation sites in GSK3β. Mass spectrometric analysis of recombinant GSK3β phosphorylated in vitro by p38 MAPK revealed two GSK3β phosphopeptides containing phosphorylation within a consensus SerPro or ThrPro motif, a phosphopeptide containing Thr^{43}, and a C-terminal peptide (384 to 403) containing the ThrPro motif at

[1]Department of Medicine/Immunobiology Program, University of Vermont, Burlington, VT 05405–0068, USA. [2]Department of Biology, University of Vermont, Burlington, VT 05405–0068, USA. [3]Department of Chemistry, University of Vermont, Burlington, VT 05405–0068, USA. [4]Department of Immunology, Roswell Park Cancer Institute, Buffalo, NY 14263, USA. [5]Program in Molecular Medicine, University of Massachusetts, Worcester, MA 01605, USA. [6]McMaster Stem Cell and Cancer Research Institute, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada.

*Present address: Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, 62210, México.
†To whom correspondence should be addressed. E-mail: mrincon@uvm.edu

Fig. 3. Inhibition of GSK3β by p38 MAPK is mediated by phosphorylation at Thr[390]. (**A**) In vitro kinase assays for recombinant p38 MAPK with catalytically inactive GSK3β and GSK3β-T[43]A mutant as substrates. (**B**) In vitro kinase assay for recombinant p38 MAPK with kinase-inactive GSK3β (WT), GSK3β-T[43]A, GSK3β-T[390]A, and GSK3β-T[43]A/T[390]A mutants as substrates. (**C**) In vitro kinase assay for recombinant active ERK with catalytically inactive GSK3β, GSK3β-T[43]A, GSK3β-T[390]A, and GSK3β-T[43]A/T[390]A mutants as substrates. (**D**) In vitro kinase assay for active GSK3β, GSK3β-T[43]A, and GSK3β-T[390]A mutants before (Con) or after incubation with activated Akt or activated p38 MAPK. GSK3β activity relative to the activity without Akt or p38 MAPK (Con) is shown. Error bars represent SD (n = 3 replicates). (**E**) In vitro GSK3β kinase reactions alone (−) or in the presence of unphosphorylated-Thr[390] (Thr[390]), phospho-Ser[9] (P-Ser[9]), or phospho-Thr[390] (P-Thr[390]) peptides, as described in (D). Error bars represent SD. (**F**) GSK3β in vitro kinase assays as in (D), using various concentrations of phospho-Thr[390], phospho-Ser[9], and unphosphorylated-Thr[390] peptides. Each point is the average of two measurements. GSM, modified glycogen synthase peptide.
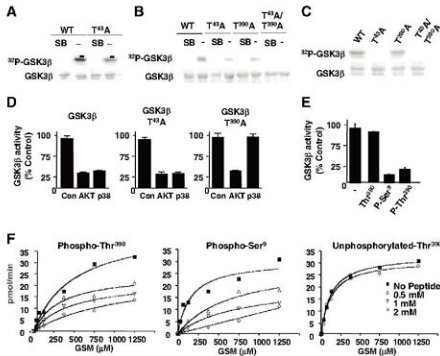
Thr[390] (corresponding to Ser389 in mouse GSK3β) (figs. S6 and S7). To confirm Thr[390] as a target of p38 MAPK in GSK3β, catalytically inactive GSK3β-T[390]A and GSK3β-T[43]A/T[390]A mutants were used as substrates for p38 MAPK in vitro. Phosphorylation of the GSK3β-T[390]A mutant by p38 MAPK was partially reduced but not abrogated (Fig. 3B), but phosphorylation of the GSK3β-T[43]A/T[390]A mutant was abrogated, indicating that these two residues are probably the targets for p38 MAPK in GSK3β. The T[43]A mutation (but not the T[390]A mutation) abrogated phosphorylation of GSK3β by ERK (Fig. 3C). Thus, Thr[390] of GSK3β appears to be specifically phosphorylated by p38 MAPK.

We examined the activity of wild-type (WT) GSK3β and GSK3β-T[43]A and GSK3β-T[390]A mutants before or after incubation with p38 MAPK or Akt. p38 MAPK inhibited both WT GSK3β and the GSK3β-T[43]A mutant (Fig. 3D), but not the GSK3β-T[390]A mutant (Fig. 3D). Akt inhibited WT GSK3β and the two mutants (Fig. 3D). p38 MAPK did not affect the activity of GSK3α (fig. S8), in which the Thr[390] residue from GSK3β is not conserved. Together, these results demonstrate that p38 MAPK–mediated phosphorylation of GSK3β at Thr[390] (but not Thr[43]) is sufficient to inhibit GSK3β activity. A peptide derived from the N terminus of GSK3β containing phospho-Ser[9] specifically inhibits GSK3β in vitro (21, 22). A phospho-Thr[390] peptide also inhibited GSK3β activity, whereas the unphosphorylated-Thr[390] peptide did not (Fig. 3E). The phospho-Thr[390] peptide inhibited GSK3β activity as efficiently as the phospho-Ser[9] peptide (Fig. 3F). Thus, phosphorylation at Thr[390] by p38 MAPK may cause an inhibition of GSK3β comparable to the phosphorylation of Ser[9] by Akt.
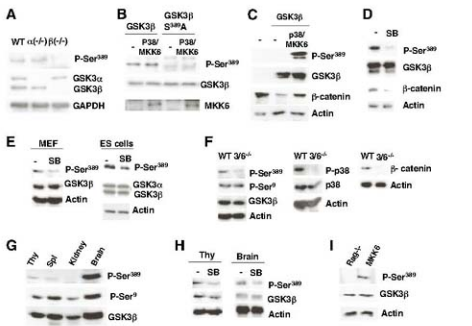
Fig. 4. Phosphorylation of endogenous GSK3β by p38 MAPK. (**A**) Western blot showing the presence of endogenous phospho-Ser[389] GSK3β (P-Ser[389]) in WT, GSK3α−/−, and GSK3β−/− ES cells. Total GSK3α, GSK3β, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were examined as controls. (**B**) Western blot showing P-Ser[389] in GSK3β−/− ES cells transfected with WT GSK3β or GSK3β-S[389]A mutant alone or with p38 MAPK and MKK6 (p38/MKK6). (**C**) Western blot showing P-Ser[389], Flag-tagged mouse GSK3β, and β-catenin in nontransfected 293T cells (−) or cells transfected with mouse GSK3β alone or in combination with p38 MAPK and MKK6 (p38/MKK6). (**D**) P-Ser[389], total GSK3β, and β-catenin in 293T cells transfected with GSK3β, p38, and MKK6 in the absence (−) or presence of SB203580. (**E**) P-Ser[389] and total GSK3β in MEF or total GSK3α and GSK3β in ES cells nontreated (−) or treated with SB203580. (**F**) P-Ser[389], P-Ser[9], total GSK3β, P-p38, total p38, and β-catenin in WT and MKK3−/−MKK6−/− (3/6−/−) MEF. (**G**) The tissue distribution of P-Ser[389], P-Ser[9], and total GSK3β. Quantification of the levels of P-Ser[389] relative to P-Ser[9] in each tissue is also shown (lower panel). Thy, thymocytes; Spl, spleen cells. (**H**) P-Ser[389] and total GSK3β in thymocytes and brain from WT mice treated in vivo with vehicle (−) or SB203580 (SB). (**I**) P-S[389] and total GSK3β in thymocytes from Rag−/− and MKK6 transgenic mice.

To demonstrate the phosphorylation of this residue in intact cells and in vivo, we generated a specific antibody (Ab) to a mouse phospho-Ser[389] GSK3β peptide. A band corresponding to GSK3β was detected with this Ab in WT and GSK3α[−/−] embryonic stem (ES) cells, but not in the GSK3β[−/−] ES cells by Western blot analysis (Fig. 4A). This specific band was also present in GSK3β[−/−] ES cells transfected with a WT GSK3β, but not with a GSK3β-S[389]A mutant (Fig. 4B). Phospho-Ser[389] GSK3β was detected in mouse GSK3β-transfected 293T cells, but only if active MKK6 was present (Fig. 4C). The presence of the phospho-Ser[389] GSK3β in these cells correlated with an increased amount of β-catenin (Fig. 4C), which is indicative of an inhibition of GSK3β activity. Ser[389]-phosphorylation was also detected in WT GSK3β, but not the GSK3β-S[389]A mutant after in vitro incubation with activated p38 MAPK (fig. S9). Phosphatase treatment of GSK3β previously incubated with activated p38 MAPK abrogated its recognition by the phospho-Ser[389] Ab (fig. S9). Together, these results show the specificity of this Ab for phospho-S[389] GSK3β and the phosphorylation of GSK3β at S[389] by p38 MAPK in vitro.

To determine whether activation of p38 MAPK was required for phosphorylation of GSK3β at Ser[389] in intact cells, we treated mouse GSK3β-transfected 293T cells with SB203580. The inhibition of p38 MAPK abrogated the phosphorylation of GSK3β (Fig. 4D). Similarly, treatment with SB203580 inhibited phosphorylation of endogenous GSK3β at Ser[389] in WT mouse embryonic fibroblasts (MEFs) and ES cells (Fig. 4E). We also examined phospho-Ser[389] abundance in MEFs deficient for the major upstream activators of p38 MAPK, MKK3, and MKK6 (23). Phospho-Ser[389] was barely detectable in MKK3[−/−] MKK6[−/−] MEFs (Fig. 4F). In contrast, the amounts of phospho-Ser[9] were comparable in WT and MKK3[−/−] MKK6[−/−] MEFs (Fig. 4F). Thus, activation of p38 MAPK appears to be required for the phosphorylation of GSK3β at Ser[389]. Inhibition of p38 MAPK by either SB203580 (Fig. 4D) or the absence of MKK3 and MKK6 (Fig. 4F) also decreased the amount of β-catenin, consistent with the possibility that p38 MAPK activation is required for repressing GSK3β activity.

We also examined phospho-Ser[389] in different mouse tissues. A high amount of phospho-S[389] was detected in the brain, and lesser amounts were detected in thymocytes and spleen cells (Fig. 4G). Phospho-Ser[389] was not detected in the kidney (Fig. 4G), liver, or heart (fig. S10). Phosphorylation of GSK3β at Ser[9] was detected in practically all of the examined tissues (Fig. 4G). Analysis of the relative abundance of phospho-S[389] and phospho-S[9] showed a predominance of the former in the brain and thymocytes (Fig. 4G), which correlated with the selective high activation of p38 MAPK in these tissues (fig. S11). Inhibition of p38 MAPK by treating animals with SB203580 reduced the levels of phospho-Ser[389]

GSK3β in both thymocytes and the brain (Fig. 4H). Analysis of phospho-Ser[389] in MKK6 and Rag[−/−] thymocytes showed that phospho-S[389] GSK3β was present selectively in MKK6 thymocytes (Fig. 4I). Together, these results support our proposal that GSK3β is phosphorylated at S[389] in vivo by p38 MAPK and that this alternative regulatory mechanism of GSK3β is tissue-specific.

To date, phosphorylation at Ser[9] by Akt is the best-characterized mechanism for the inhibition of GSK3β activity. However, knockin mice in which Ser[9] was replaced by Ala have only a subtle defect related to insulin regulation of glycogen synthase in their skeletal muscle tissue (24), indicating that alternative mechanisms may be involved in the negative regulation of GSK3β for certain functions. We propose that the phosphorylation of GSK3β at S[389] by p38 MAPK may be one such mechanism. Conditions that promote the accumulation of β-catenin in certain scenarios; thus, the activation of the p38 MAPK pathway could be an alternative mechanism to regulate β-catenin/T cell factor signaling (and, potentially, cell survival) through the inactivation of GSK3β.

References and Notes
1. T. Zarubin, J. Han, Cell Res. **15**, 11 (2005).
2. A. Mikhailov, M. Shinohara, C. L. Rieder, Cell Cycle **4**, 57 (2005).
3. Q. B. She, N. Chen, Z. Dong, J. Biol. Chem. **275**, 20444 (2000).
4. T. Kurosu et al., Apoptosis **10**, 1111 (2005).
5. H. C. Reinhardt, A. S. Aslanian, J. A. Lees, M. B. Yaffe, Cancer Cell **11**, 175 (2007).
6. N. I. Dmitrieva, D. V. Bulavin, A. J. Fornace Jr, M. B. Burg, Proc. Natl. Acad. Sci. U.S.A. **99**, 184 (2002).
7. N. L. Diehl et al., J. Exp. Med. **191**, 321 (2000).
8. G. Pedraza-Alva et al., EMBO J. **25**, 763 (2006).
9. V. Ioannidis, F. Beermann, H. Clevers, W. Held, Nat. Immunol. **2**, 691 (2001).
10. F. Gounari et al., Nat. Immunol. **2**, 863 (2001).
11. B. Hoffman, A. Amanullah, M. Shafarenko, D. A. Liebermann, Oncogene **21**, 3414 (2002).
12. Materials and methods are available as supporting material on Science Online.
13. T. C. He et al., Science **281**, 1509 (1998).
14. M. Fibali, N. Cheng, D. Abbott, V. Leontiev, J. F. Engelhardt, J. Biol. Chem. **277**, 33398 (2002).
15. C. Liu et al., Cell **108**, 837 (2002).
16. S. Salabhor, J. R. Woodgett, J. Clin. Pathol. **58**, 225 (2005).
17. D. A. Cross, D. R. Alessi, P. Cohen, M. Andjelkovich, B. A. Hemmings, Nature **378**, 785 (1995).
18. B. W. Doble, S. Patel, G. A. Wood, L. K. Kockeritz, J. R. Woodgett, Dev. Cell **12**, 957 (2007).
19. B. W. Doble, J. R. Woodgett, J. Cell Sci. **116**, 1175 (2003).
20. Q. Ding et al., Mol. Cell **19**, 159 (2005).
21. R. Dajani et al., Cell **105**, 721 (2001).
22. S. Frame, P. Cohen, Biochem. J. **359**, 1 (2001).
23. D. Brancho et al., Genes Dev. **17**, 1969 (2003).
24. E. J. McManus et al., EMBO J. **24**, 1571 (2005).
25. We thank C. Charland for flow cytometry analysis and cell sorting, T. Hunter and the DNA Sequencing Facility for sequencing, and the Univ. of Vermont Protein Core Facility for peptide synthesis. This work was supported by NIH grants R01 AI051454 and National Center for Research Resources (NCRR) grant P20 RR15557 (M.R.), NCRR grants P20 RR021905 and P20 RR16462 (D.E.M.), and Canada Research Chairs and Canadian Institutes of Health Research grant MOP-85057 (B.D.J.).

Asymmetric Tethering of Flat and Curved Lipid Membranes by a Golgin

Guillaume Drin,[1] Vincent Morello,[1] Jean-François Casella,[1] Pierre Gounon,[2] Bruno Antonny[1]*

Golgins, long stringlike proteins, tether cisternae and transport vesicles at the Golgi apparatus. We examined the attachment of golgin GMAP-210 to lipid membranes. GMAP-210 connected highly curved liposomes to flatter ones. This asymmetric tethering relied on motifs that sensed membrane curvature both in the N terminus of GMAP-210 and in ArfGAP1, which controlled the interaction of the C terminus of GMAP-210 with the small guanine nucleotide–binding protein Arf1. Because membrane curvature constantly changes during vesicular trafficking, this mode of tethering suggests a way to maintain the Golgi architecture without compromising membrane flow.

The Golgi apparatus has a stable architecture despite the intense flux of membrane that passes through it (1, 2). Golgins, which probably correspond to molecular strings observed by electron microscopy (3), contribute to this architecture by tethering membranes thanks to their coiled-coil structure, up to 200 nm in length (4). Some golgins link specific transport vesicles to cisternae and may restrict their diffusion (5). But because vesicles bud and fuse

within minutes there must be regulatory mechanisms to promote and disrupt these links within the same time scale. We studied human golgin GMAP-210. This protein is located at the cis

[1]Institut de Pharmacologie Moléculaire et Cellulaire, Université de Nice Sophia Antipolis and CNRS, 660 route des lucioles, 06560 Valbonne, France. [2]Centre Commun de Microscopie Appliquée, Parc Valrose, 06108 Nice Cedex 2, France.

*To whom correspondence should be addressed. E-mail: antonny@ipmc.cnrs.fr

Golgi, and its overexpression induces the formation of clusters of small vesicles (radius, $R \approx$ 30 nm) at the expense of the Golgi (6, 7).

The first 38 residues of GMAP-210 form an ALPS (amphipathic lipid-packing sensor) motif, a lipid-binding module that is remarkably sensitive to membrane curvature (8). Soluble and unfolded in the presence of weakly curved liposomes ($R > 100$ nm), ALPS motifs form an amphipathic α helix at the surface of small liposomes ($R < 50$ nm) (8, 9). Through this motif, GMAP-210 could trap small transport vesicles. On its C terminus, GMAP-210 is predicted to contain a GRAB (GRIP-related Arf binding) domain that may interact with the small guanine nucleotide–

binding protein Arf1 in the guanosine 5′-triphosphate (GTP) state (10). We reasoned that this putative interaction could be stable only on flat membranes because ArfGAP1, a guanosine triphosphatase (GTPase) activating protein for Arf1 at the Golgi, contains two ALPS motifs, making its activity exquisitely dependent on membrane curvature (9, 11, 12). Thus, the presence of ALPS motifs both in GMAP-210 and in ArfGAP1 suggests an asymmetric mode of tethering between flat and curved membranes (Fig. 1A). In vivo, this tethering would be ideally suited to capture vesicles in the vicinity of flat cisternae. Moreover tethering should be disrupted as soon as one of the two membranes lost its

identity: that is, when the vesicle becomes flat by membrane fusion or when the cisternae becomes curved by budding. Thus, GMAP-210 could readily recycle such as to be always properly orientated.

Interaction between the GRAB domain of GMAP-210 and a soluble form of Arf1GTP has not been observed (10). Because Arf1 interacts with its partners at the surface of lipid membranes, we reassessed its interaction with GMAP-210 in the presence of liposomes. Two truncated forms of GMAP-210 (GMAP$_{C\text{-short}}$ contained amino acids 1597 to 1830 and GMAP$_{C\text{-long}}$ included amino acids 1597 to 1843) were used (Fig. 1B and table S1). Both included part of the coiled-coil domain and the GRAB domain, but the longer construct contained a short predicted amphipathic helix downstream from the GRAB domain. These constructs were incubated with liposomes in the presence of Arf1, and liposome-bound proteins were recovered by flotation. Arf1GTP but not Arf1GDP efficiently recruited GMAP$_{C\text{-long}}$ to the liposomes (Fig. 1C). No recruitment by Arf1GTP was observed for GMAP$_{C\text{-short}}$ or when GMAP$_{C\text{-long}}$ carried the Leu$^{1783} \rightarrow$ Ala^{1783} (L1783A) mutation (Fig. 1C and fig. S1). This mutation eliminates the Golgi localization of the C-terminal region of GMAP-210, and the cognate residue in the GRAB domain is critical for the interaction with Arf1GTP (10). Thus, in the same way as the GRIP domain (13, 14), the GRAB domain may interact simultaneously with the lipid membrane through its amphipathic helix and with Arf1GTP (Fig. 1A). To test this, we attached a membrane-sensitive fluorescent probe (NBD) to this helix via a cysteine mutation (fig. S2, A to C). Guanosine diphosphate (GDP)–to–GTP exchange on Arf1 promoted an increase and a blue shift in the fluorescence of ^{NBD}GMAP$_{C\text{-long}}$, suggesting that the amphipathic helix of the GRAB domain contacted the liposome surface (Fig. 1D and fig. S3, A to D). The kinetics of the NBD fluorescence change matched the time course of Arf1 activation, and its amplitude varied with Arf1 concentration in a manner suggesting a stoichiometric interaction between the two proteins (fig. S3B).

Two GTPase activating proteins for Arf1, ArfGAP1, and ArfGAP3 (Gcs1p and Glo3p in yeast) are localized at the cis Golgi (15) and are candidates for promoting GTP hydrolysis in the Arf1GTP-GRAB domain complex. To test this, we took advantage of the fluorescence signal associated with the translocation of ^{NBD}GMAP$_{C\text{-long}}$. ArfGAP1 reversed the NBD signal within minutes, whereas ArfGAP3 had almost no effect (Fig. 1D). This observation fits with the idea that ArfGAP1/Gcs1p can act on several Arf/Arf-effector complexes at the Golgi, including complexes with golgins, whereas ArfGAP3/Glo3p is more specific for the Arf1-coatomer complex (16–18). Because the activity of ArfGAP1 is strongly dependent on membrane curvature owing to its ALPS motifs (9, 12), we asked whether ArfGAP1
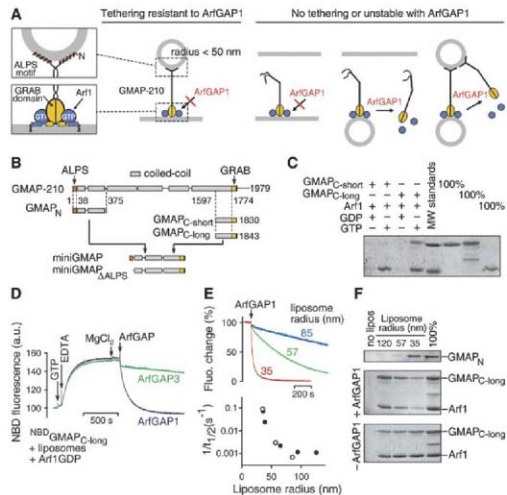
Fig. 1. Membrane attachment of the two ends of GMAP-210. (A) With its N-terminal ALPS motif and its C-terminal GRAB domain, GMAP-210 could connect small vesicles to membranes containing Arf1GTP. The former interaction is stable only on curved membranes (8), whereas the latter should be stable only on flat membranes because ArfGAP1 contains ALPS motifs and is very active on curved membranes. (B) Domain organization of GMAP-210 and scheme of the constructs (24). (C) GMAP$_{C\text{-short}}$ or GMAP$_{C\text{-long}}$ (0.5 μM) was incubated with liposomes (0.5 mM lipids, $R = 115 \pm 51$ nm), with Arf1 (1 μM), and with GDP or GTP (66 μM) at 1 μM free Mg^{2+}. The liposomes were recovered by flotation and bound proteins were analyzed by SDS–polyacrylamide gel electrophoresis. (D) NBD fluorescence assay. The cuvette contained liposomes (0.2 mM lipids, $R = 42 \pm 8$ nm), ^{NBD}GMAP$_{C\text{-long}}$ (0.125 μM), and Arf1GDP (0.75 μM). When indicated, GTP (10 μM) was added, and nucleotide exchange was promoted by lowering temporally the concentration of free Mg^{2+}. Thereafter ArfGAP1 or ArfGAP3 (0.1 μM) was added. (E) Effect of membrane curvature. (Top) Same as in (D) with liposomes of different radius. (Bottom) Rate of ^{NBD}GMAP$_{C\text{-long}}$ dissociation versus liposome radius as determined from two independent experiments. (F) Opposite effect of membrane curvature on the N and C termini of GMAP-210. GMAP$_{C\text{-long}}$ was incubated with the liposomes and with Arf1 and GTP as in (C) followed by 3-min incubation with 0.1 μM ArfGAP1. GMAP$_N$ was simply mixed with the liposomes.

retained this feature when it promoted GTP hydrolysis in the Arf1GTP-GMAP$_{C-long}$ complex. Reducing liposome size from 100 to 35 nm accelerated 100-fold the dissociation of the complex triggered by Arf1GTP (Fig. 1E). This response resembles that observed on liposomes covered solely by Arf1GTP or by Arf1 GTP in complex with coatomer (11). Thus, if the Arf1GTP-GRAB domain complex is not sensitive to membrane curvature per se, the control of its attachment by ArfGAP1 creates sensitivity to membrane curvature opposite to that of the N terminus. In liposome binding experiments, the N-terminal region of GMAP-210 (GMAP$_N$) included amino acids 1 to 375) bound preferentially to small liposomes, whereas the complex between the C-terminal region (GMAP$_{C-long}$) and Arf1 GTP was stable, in the presence of ArfGAP1, only on large liposomes (Fig. 1F).

Having defined simple rules for the membrane attachment of the two ends of GMAP-210, we wished to reconstitute its tethering activity on liposomes. Because full-length GMAP-210 was difficult to express in E. coli and poorly soluble, we used a shorter construct, miniGMAP, made by fusing GMAP$_N$ and GMAP$_{C-long}$ (Fig. 1B). MiniGMAP contained the ALPS motif, one-third of the coiled-coil region, and the GRAB domain with its amphipathic helix. When over-expressed in HeLa cells, this construct affected the Golgi morphology in a manner similar to that of the full-length form (figs. S4 and S5). We assessed the effect of miniGMAP on a homogeneous population of liposomes. Dynamic light scattering (DLS) and electron microscopy (EM) were used to detect liposome aggregation. MiniGMAP caused small liposomes ($R = 32 \pm 9$ nm) covered with Arf1GTP to assemble in large aggregates (0.5 to 1 μm) within minutes (Fig. 2, A and B, and fig. S6). In contrast, almost no aggregation was observed for liposomes devoid of Arf1GTP or when miniGMAP lacked the ALPS motif (Fig. 2A and fig. S7). Liposome aggregation also diminished when vesicle size increased (see below). Thus, the tethering activity of GMAP-210 relies both on its ALPS motif and on the interaction of its GRAB domain with Arf1 GTP. Tethering was very efficient: Liposomes aggregated with only 25 nM miniGMAP (fig. S6), a concentration 10- to 100-fold lower than those in other tethering reactions reconstituted with proteins and liposomes (19, 20). This corresponds to 10 to 15

copies of miniGMAP per liposome (table S2), a density similar to what is used to artificially dock liposomes through complementary DNA molecules (21).

Next, we conducted experiments in which two populations of liposomes of defined size were mixed: one covered with Arf1GTP and one devoid of Arf1 (Fig. 2C). Both were used at the same concentration of accessible lipids. Shortly after liposome mixing, miniGMAP was added, and aggregation was followed. Lastly, ArfGAP1 was added to test the resistance of the liposome aggregates. Strong aggregation was observed for two mixtures: those containing small naked liposomes and small liposomes covered with Arf1GTP and those containing small naked liposomes and large liposomes covered with Arf1GTP (Fig. 2C). In contrast, the aggregation signal was much weaker when both liposome populations were of large size (fig. S8). Thus for membrane aggregation to occur, the presence of highly curved membranes and of membrane-bound Arf1GTP is required (Fig. 2A), but these two determinants do not need to be on the same liposome.

The asymmetric and the symmetric tethering geometries differed in their sensitivity to ArfGAP1.
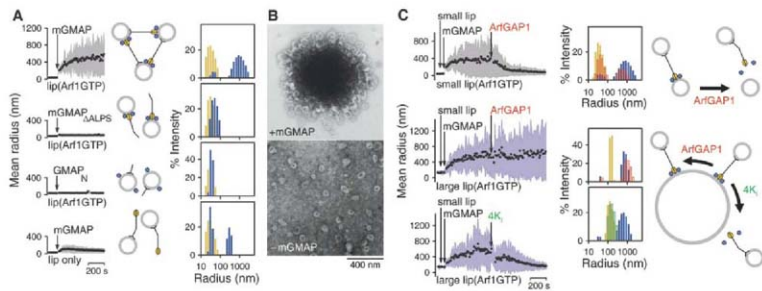
Fig. 2. Liposome tethering induced by miniGMAP. (**A**) Small liposomes ($R = 32 \pm 9$ nm, 50 μM accessible lipids) covered or not with Arf1GTP (0.25 μM) were mixed with miniGMAP, miniGMAP$_{\Delta ALPS}$, or GMAP$_N$ (0.125 μM). Liposome aggregation was followed by DLS. (Left) Mean radius and polydispersity (shaded area) over time. (Right) Size distribution before (yellow bars) and after (blue bars) the reaction. (**B**) Electron micrographs of negatively stained small liposomes ($R = 38 \pm 9$ nm) covered with Arf1GTP and incubated or not with miniGMAP as in (A). (**C**) Two populations of liposomes (small, $R = 36 \pm 7$ nm; large, $R = 143 \pm 45$ nm; 25 μM accessible lipids each) and either covered or not with Arf1GTP (0.25 μM) were mixed. Then 0.125 μM miniGMAP was quickly added, and liposome aggregation was followed by DLS. Thereafter, ArfGAP1 or the 4K$_i$ mutant was added at 0.25 μM. Yellow bars indicate initial size distribution; blue bars, size distribution after aggregation; and red or green bars, final size distribution after ArfGAP1 or 4K$_i$, respectively, addition. (**D**) Typical assembly of large (l) liposomes ($R = 144 \pm 55$ nm; 25 μM accessible lipids) covered with 0.125 μM Arf1GTP after incubation with 62.5 nM miniGMAP and with small (s) naked liposomes ($R = 41 \pm 16$ nm; 50 μM accessible lipids). See figs. S6 to S9 for supplementary data and gallery of electron micrographs.

The majority of the aggregates formed in mixture containing only small liposomes disassembled within minutes upon ArfGAP1 addition, whereas those formed by large liposomes covered with Arf1GTP and small naked liposomes were resistant (Fig. 2C). In the latter case, we observed disassembly when we used a mutant of ArfGAP1 ($4K_t$) that displays more avidity to flat membranes than the wild-type form owing to specific mutations in its ALPS motif (8). Thus, ArfGAP1 through its ALPS motif is capable of selectively disrupting miniGMAP-induced tethering according to the curvature of the membrane on which Arf1GTP anchors the GRAB domain. The fact that ArfGAP1 or the $4K_t$ mutant can reverse liposome aggregation suggests that no substantial membrane fusion occurred during tethering. EM analysis of incubations conducted with small naked liposomes and large liposomes supplemented with Arf1GTP revealed the formation of clusters containing a few large liposomes and many small liposomes, the latter forming a kind of cement around the larger ones, suggesting that large liposomes contact preferentially small liposomes and vice versa (Fig. 2D and fig. S9). If membrane tethering was random, we should have observed direct contacts between large liposomes as well as clusters of small liposomes such as those visualized previously (Fig. 2B). Thus, in the presence of ArfGAP1, GMAP-210 forms an asymmetric tether that can stably connect a curved membrane to a flat one displaying Arf1GTP but not other geometrical combinations (Fig. 1A).

Next we established a system suitable for light microscopy. We mixed giant liposomes (tens of micrometers in size) labeled with a green fluorophore with small liposomes labeled with a red fluorophore. The former were visible under the microscope, whereas the latter gave a red fluorescence background. Upon attachment of small liposomes, the contour of the giant liposomes became red (fig. S10). Arf1 was allowed to undergo cycles of GTP binding and hydrolysis on the two populations of liposomes, caused by the addition of ArfGAP1 and of the phosphatidylinositol 4,5-bisphosphate (PIP_2)-dependent guanine nucleotide exchange factor Arno (22). Because both large and small liposomes contained PIP_2, Arno-catalyzed GDP/GTP exchange on Arf1 occurred on the two populations of liposomes, whereas ArfGAP1-catalyzed GTP hydrolysis in Arf1 occurred preferentially on the small liposomes owing to their strong curvature. Thus, with both Arno and ArfGAP1 present, Arf1GTP should be found at steady state preferentially on the giant liposomes. Under these conditions, miniGMAP caused the large liposomes to be surrounded by red fluorescence (Fig. 3, A to C, and fig. S11). In contrast, when ArfGAP1 was absent, we observed numerous red spots of various size and rarely connected to the giant liposomes (Fig. 3D). Thus, ArfGAP1 helps to organize GMAP-210 tethering by pre-

venting the formation of symmetric assemblies between small liposomes, thereby favoring asymmetric assemblies between large and small liposomes.

Multiple tethering events involving different membranes and several long coiled-coil proteins (e.g., p115) occur at the interface between the endoplasmic reticulum and the cis Golgi (4). The minimal model presented here seems adapted to the capture of small transport vesicles at this region, but additional interactions with protein coats, Rabs, and the cytoskeleton may impose a more specific role to GMAP-210 (4, 23). Nevertheless, a reversible tethering mechanism based on the detection of membrane curvature is straightforward because curvature is a good index for the completion of budding and fusion events. By permitting transient interactions between membranes that are continuously remodeled, GMAP-210 may contribute to the self-organizing properties of the Golgi.

References and Notes
1. T. Misteli, *J. Cell Biol.* **155**, 181 (2001).
2. N. Altan-Bonnet, R. Sougrat, J. Lippincott-Schwartz, *Curr. Opin. Cell Biol.* **16**, 364 (2004).
3. L. Orci, A. Perrelet, J. E. Rothman, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2279 (1998).
4. B. Short, A. Haas, F. A. Barr, *Biochim. Biophys. Acta* **1744**, 383 (2005).
5. J. Malsam, A. Satoh, L. Pelletier, G. Warren, *Science* **307**, 1095 (2005).
6. R. M. Rios *et al.*, *J. Cell Biol.* **125**, 997 (1994).
7. K. Pernet-Gallay *et al.*, *Traffic* **3**, 822 (2002).
8. G. Drin *et al.*, *Nat. Struct. Mol. Biol.* **14**, 138 (2007).
9. J. Bigay, J. F. Casella, G. Drin, B. Mesmin, B. Antonny, *EMBO J.* **24**, 2244 (2005).
10. A. K. Gillingham, A. H. Tong, C. Boone, S. Munro, *J. Cell Biol.* **167**, 281 (2004).
11. J. Bigay, P. Gounon, S. Robineau, B. Antonny, *Nature* **426**, 563 (2003).
12. B. Mesmin *et al.*, *Biochemistry* **46**, 1779 (2007).
13. B. Panic, O. Perisic, D. B. Veprintsev, R. L. Williams, S. Munro, *Mol. Cell* **12**, 863 (2003).
14. L. Lu, G. Tai, M. Wu, H. Song, W. Hong, *Traffic* **7**, 1399 (2006).
15. H. Inoue, P. A. Randazzo, *Traffic* **8**, 1465 (2007).
16. P. P. Poon *et al.*, *EMBO J.* **18**, 555 (1999).
17. Y. W. Liu, C. F. Huang, K. B. Huang, F. J. Lee, *Mol. Biol. Cell* **16**, 4024 (2005).
18. A. Fagster, G. Frigerio, M. Dale, R. Duden, *EMBO J.* **19**, 3905 (2000).
19. H. Nakatogawa, Y. Ichimura, Y. Ohsumi, *Cell* **130**, 165 (2007).
20. D. Arac *et al.*, *Nat. Struct. Mol. Biol.* **13**, 209 (2006).
21. Y. H. Chan, P. Lenz, S. G. Boxer, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18913 (2007).
22. P. Chardin *et al.*, *Nature* **384**, 481 (1996).
23. R. M. Rios, A. Sanchís, A. M. Tassin, C. Fedriani, M. Bornens, *Cell* **118**, 323 (2004).
24. Materials and methods are available on Science Online.
25. We thank R. Rios, J. Bigay, and D. Stalder for reagents; N. Leroudier, S. Scarzello, and F. Braud for technical assistance; and C. Jackson and B. Goud for comments on the manuscript. This work was supported by the Agence Nationale de la Recherche (programme blanc) and by the CNRS (programme interface).
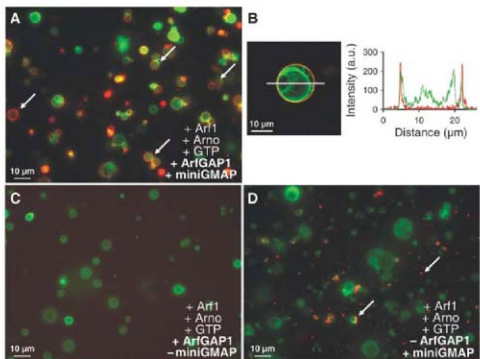
Fig. 3. Self-organization of membrane tethering by miniGMAP when the GTPase cycle of Arf1 is controlled by Arno and ArfGAP1. The sample contains giant liposomes (green) and small liposomes (red; $R = 42 \pm 10$ nM), all with the same composition (Golgi mix plus 4% PIP_2). Arrows indicate typical liposome assemblies. (**A** and **B**) The liposome mixture was supplemented with Arf1 (380 nM), Arf1 (150 μM), Arno (20 nM), ArfGAP1 (185 nM), and miniGMAP (62.5 nM). a.u., arbitrary units. (**C** and **D**) MiniGMAP or ArfGAP1 was omitted. The liposomes were visualized by epifluorescence microscopy [(A), (C), and (D)] or by confocal microscopy (B). Other confocal images are shown in fig. S11.

Innate Immune Activation Through Nalp3 Inflammasome Sensing of Asbestos and Silica

Catherine Dostert,[1] Virginie Pétrilli,[1] Robin Van Bruggen,[2] Chad Steele,[3] Brooke T. Mossman,[4] Jürg Tschopp[1]*

The inhalation of airborne pollutants, such as asbestos or silica, is linked to inflammation of the lung, fibrosis, and lung cancer. How the presence of pathogenic dust is recognized and how chronic inflammatory diseases are triggered are poorly understood. Here, we show that asbestos and silica are sensed by the Nalp3 inflammasome, whose subsequent activation leads to interleukin-1β secretion. Inflammasome activation is triggered by reactive oxygen species, which are generated by a NADPH oxidase upon particle phagocytosis. (NADPH is the reduced form of nicotinamide adenine dinucleotide phosphate.) In a model of asbestos inhalation, Nalp3$^{-/-}$ mice showed diminished recruitment of inflammatory cells to the lungs, paralleled by lower cytokine production. Our findings implicate the Nalp3 inflammasome in particulate matter–related pulmonary diseases and support its role as a major proinflammatory "danger" receptor.

Inhalation of asbestos or silica in occupational exposures can result in pulmonary fibrosis (asbestosis, silicosis), and lung cancer, especially in smokers (1). Asbestos fibers have been associated with development of malignant mesotheliomas after environmental exposures (2). The mechanisms of injury to cells of the lung and pleura and disease development by these pathogenic particulates are unclear (3, 4). Inflammation is a hallmark of exposure to asbestos or silica and is observed both in animal models and in the lungs of patients with asbestos-related lung disease (1, 3). Bronchial epithelial cells and alveolar macrophages are in prolonged contact with the inhaled particulates when clearing them from the lung and can initiate and sustain inflammatory responses. Likewise, inflammatory responses are observed after exposure to other particulates, such as those found in diesel exhaust (5).

The reported proinflammatory activities of asbestos and silica prompted us to investigate the potential of these particulates in activating the secretion of the proinflammatory cytokine interleukin-1β (IL-1β) in human macrophages (6). We indeed observed that the macrophage-like cell line THP1 secreted mature IL-1β on stimulation with asbestos or silica (Fig. 1A). In contrast, the addition of diesel exhaust particles (DEPs) or cigarette smoke extract (CSE) did not result in the production of detectable IL-1β when measured after 6 hours of exposure (Fig. 1A). The secretion of mature IL-1β induced by asbestos was comparable to amounts observed with monosodium urate crystals (MSU), known to be a potent activator of IL-1β processing (7) (Fig. 1, A and B). Inert particles, such as polystyrene beads, did not activate IL-1β maturation (Fig. 1B). Primary

human monocyte-derived macrophages also produced mature IL-1β after stimulation with asbestos or silica (Fig. 1C), whereas DEP exposure did not result in IL-1β secretion.

Mature IL-1β is produced by cleavage of the inactive pro–IL-1β precursor by caspase-1, which is activated within a large multiprotein complex, termed the inflammasome (8). The Nalp3 inflammasome, composed of the Nod-like receptor (NLR) protein Nalp3 (also called Cryopyrin or NLRP3), Cardinal, the adaptor ASC (apoptosis-associated speck-like protein containing a C-terminal caspase recruitment domain), and caspase-1, is implicated in the production of mature IL-1β in response to a variety of signals. For example, the presence of bacteria is recognized through binding of the pathogen-associated molecular pattern (PAMP) muramyldipeptide (MDP) or via the action of bacterial toxins (9, 10). In addition to PAMPs, Nalp3 has the extraordinary capacity of sensing endogenous stress-associated danger signals (DAMPs) such as adenosine triphosphate (ATP) or MSU (7). Because both MSU and asbestos are crystalline structures, we wondered whether production of IL-1β by asbestos would also occur through the Nalp3 inflammasome. To clarify this, we determined whether the particulates activate caspase-1 in THP1 cells and, indeed, found processing of caspase-1 into the p10 fragment (fig. S1). We then measured IL-1β maturation in cells in which the different inflammasome components were down-regulated. Highly reduced secretion of mature IL-1β was observed in cells from which Nalp3, ASC, and caspase-1 had each been removed (knocked down), similar patterns were observed with MSU or the bacterial toxin in nigericin (Fig. 2A). The residual IL-1β secretion observed in these cells is most likely due to incomplete shutdown of the corresponding proteins (fig. S2). In contrast, Toll-like receptors (TLRs) or IL-1 receptors do not seem to be essential for asbestos signaling, because the knockdown or knock-out of MyD88, an intracellular adaptor protein mediating TLR and IL-1 receptor signaling, had little or no effect on IL-1β secretion and caspase-1 activation (Fig. 2A and fig. S3). To

[1]Department of Biochemistry, University of Lausanne, Chemin des Boveresses 155, 1066 Epalinges, Switzerland. [2]Department of Blood Cell Research, Sanquin Research and Landsteiner Laboratory, Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands. [3]Department of Medicine, Division of Pulmonary, Allergy, and Critical Care Medicine, University of Alabama at Birmingham School of Medicine, Birmingham, AL 35294, USA. [4]Department of Pathology, University of Vermont College of Medicine, Burlington, VT 05405, USA.

*To whom correspondence should be addressed. E-mail: jurg.tschopp@unil.ch
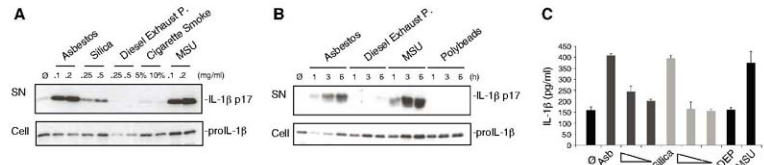
Fig. 1. Asbestos and silica activate IL-1β secretion in human macrophages. (A) THP1 cells were stimulated for 6 hours with the indicated amounts (mg/ml media) of asbestos, silica, DEPs, CSE (% in solution), or MSU. (B) THP1 cells were stimulated with 0.2 mg/ml asbestos, 0.5 mg/ml DEPs, or 0.2 mg/ml MSU for the indicated times. Media supernatants (SN) were analyzed for the presence of mature IL-1β and cell extracts (Cell), for the presence of pro–IL-1β by Western blotting. (C) LPS-primed primary human monocyte-derived macrophages (M-CSF) were stimulated for 6 hours with asbestos (0.2, 0.1, or 0.05 mg/ml); silica (0.5, 0.2, or 0.1 mg/ml); DEP (0.2 mg/ml); and MSU (0.2 mg/ml) and were analyzed by enzyme-linked immunosorbent assay (ELISA) for IL-1β released into the media. Values are means ± SEM.

further analyze the requirement for the Nalp3 inflammasome in particulate-induced IL-1β processing, we used primary murine bone marrow–derived macrophages primed with lipopolysaccharide (LPS) to induce pro–IL-1β synthesis and then stimulated them with asbestos, silica, or MSU (Fig. 2B). Caspase-1 was no longer cleaved and secreted in the presence of the pan-caspase inhibitor zVAD or in macrophages from Nalp3- and ASC-deficient mice, and as a consequence, mature IL-1β processing and secretion were abolished (Fig. 2B). By contrast, IL-1β production was independent of the Ipaf inflammasome, which has been shown to activate caspase-1 and to produce IL-1β in response to flagellin from *Salmonella typhimurium* and *Legionella pneumophila* (fig. S4) (11, 12).

The mechanism of Nalp3 inflammasome activation by PAMPs or DAMPs is still poorly understood. There are at least two possible hypotheses: the activating molecules could directly interact with the LRR domain of Nalp3 after entering the cell, or they could modify one or more membrane-associated proteins, which then trigger a signaling cascade leading to Nalp3 activation. In favor of the latter assumption is the recent finding that potassium (K^+) efflux, lowering intracellular K^+ levels, is a requirement for Nalp3 inflammasome activation triggered by all known activators including MSU (13–15). We inhibited K^+ efflux by adding high concentrations (130 mM) of KCl to the culture medium of THP1 cells and found that asbestos-induced IL-1β production was blocked (Fig. 3A).

In order to delineate the asbestos-induced signaling pathway leading to inflammasome activation in more detail, we started with the question of whether asbestos fibers needed to be endocytosed. Phagocytic cells can endocytose small particles, whereas bigger crystals or fibers are subject to so-called "frustrated" phagocytosis (16), and remain trapped at the surface (Fig. 3B). To test the importance of the endocytic process, we treated THP1 cells with cytochalasin D, which disrupts actin filaments. Cytochalasin D inhibited not only secretion of mature IL-1β by asbestos (Fig. 3C), but also by the particulate MSU, whereas for noncrystalline Nalp3 activators such as R837 and nigericin, IL-1β secretion was not affected (Fig. 3C). Because the mature form of IL-1β did not accumulate inside the cell (Fig. 3C), we conclude that the actin cytoskeleton is necessary for the attempt to phagocytose but does not drive mature IL-1β secretion. The same effect was achieved by treating the cells with another inhibitor of actin polymerization, latrunculin A (fig. S5).

Asbestos fibers have been shown to participate in redox reactions to generate reactive oxygen species (ROS) that are widely linked to signaling pathways and cause inflammation and carcinogenesis [reviewed in (17)]. The generation of ROS correlates with toxicity and pathogenicity of different types of fibers. Recent data also suggest a role for ROS in activating caspase-1 on ATP treatment (18) and in Nalp3 inflammasome activation by MSU and R837 (13). It is, therefore, possible that asbestos particles, through activation of the reduced form of

nicotinamide adenine dinucleotide phosphate (NADPH) oxidase triggered by "frustrated" phagocytosis, generate ROS, which, in turn, contribute to inflammasome activation. In order to test this hypothesis, we investigated whether asbestos led to ROS generation under our experimental conditions, which was the case (fig. S6). We then treated THP1 cells with the ROS inhibitors *N*-acetyl-L-cysteine (NAC) or (2R,4R)-4-aminopyrrolidine-2,4-dicarboxylate (APDC) to determine whether the redox status of the cell affected inflammasome activation. IL-1β production was indeed impaired in response to asbestos, MSU, and ATP when using these inhibitors (Fig. 3D), which suggested that ROS production was a necessary step in inflammasome activation. Crocidolite asbestos fibers appear to be particularly potent ROS producers because of the presence of iron on the fiber surface. The iron is thought to play an important role in the generation of free radicals (3). We observed that iron chelation of asbestos fibers by deferoxamine strongly reduced IL-1β maturation, although it had no effect in response to MSU crystals (Fig. 3E). However, uricase treatment blocked the activity of MSU but not of asbestos.

Our model predicts that inflammasome-activating ROS are generated by a NADPH oxidase known to be assembled and activated on phagocytosis of microbes. In support of this, IL-1β production was inhibited by the NADPH oxidase inhibitors diphenyleneiodonium (DPI) and apocynin, but not by rotenone (an inhibitor of mitochondrial complex I) or TTFA (an inhibitor of mitochondrial complex II) (fig. S7). To further corroborate NADPH oxidase as a source of ROS production, we knocked down the common NADPH oxidase subunit p22^{phox} and found highly diminished IL-1β secretion (Fig. 3F). If ROS play a central role in NALP3 inflammasome activation, ROS detoxifying proteins, such as thioredoxin (TRX), should be a regulating role in inflammasome activation. Indeed, when asbestos and MSU were added to macrophages in which TRX was down-regulated by short hairpin RNA (shRNA), increased IL-1β secretion was observed (Fig. 3G).

Models of asbestos inhalation have revealed that the airway epithelium and alveolar macrophages play important roles in cell proliferation, inflammation, and fibrogenesis triggered by the asbestos fibers [reviewed in (1)]. To investigate the in vivo significance of the Nalp3 inflammasome in asbestos-induced inflammation, Nalp3$^{-/-}$ and Nalp3$^{+/+}$ littermate mice were exposed for 9 days to chrysotile asbestos and markers of injury, inflammation, and cytokine production were analyzed on day 10. As previously shown (19, 20), asbestos-exposed mice exhibited increased total cell numbers in bronchoalveolar lavage fluid (BALF) compared with air-exposed mice. However, significantly fewer cells were recruited to the lungs of Nalp3$^{-/-}$ mice after exposure to asbestos (Fig. 4A). Notably, lymphocyte, eosinophil, and neutrophil infiltration was reduced in
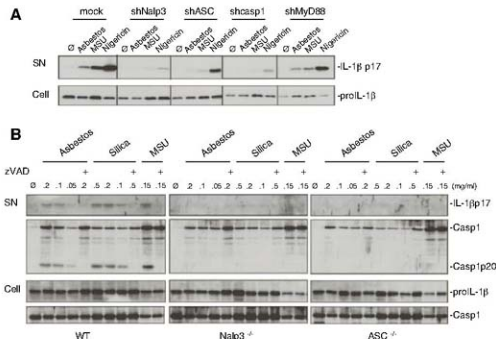
Fig. 2. Asbestos-induced IL-1β secretion is dependent on Nalp3 inflammasome. **(A)** THP1 cells stably expressing shRNA against the indicated target genes were stimulated for 6 hours with 0.2 mg/ml asbestos, 0.2 mg/ml MSU, and 3.4 μM nigericin. Media supernatants (SN) and cell extracts (Cell) were analyzed by Western blotting as indicated in the text. **(B)** LPS-primed, murine bone marrow–derived macrophages from wild-type (WT), Nalp3- or ASC-deficient mice were stimulated for 6 hours with the indicated amounts (mg/ml) of asbestos, silica, and MSU, in the presence of 20 μM zVAD where indicated, and were analyzed for IL-1β secretion and activation of caspase-1 by Western blotting.
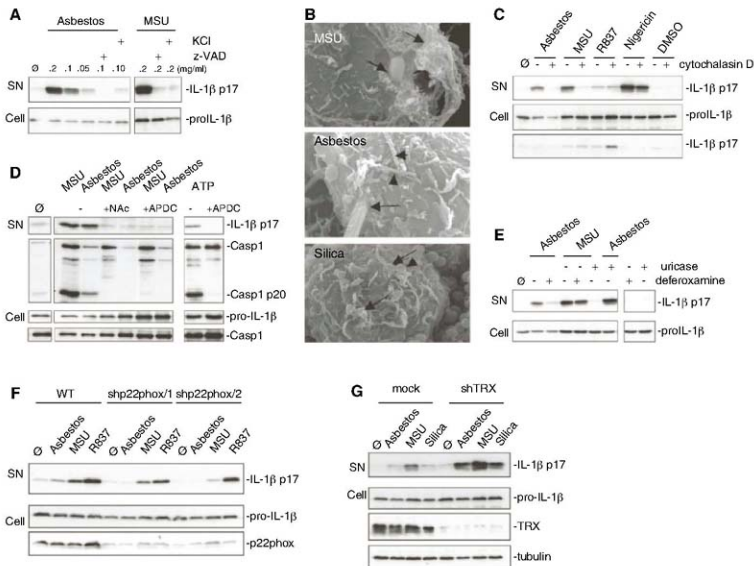
Fig. 3. Nalp3 inflammasome activation after asbestos stimulation is dependent on endocytosis and ROS production. (**A**) THP1 cells were stimulated with the indicated amounts of asbestos or MSU, in the presence of 20 μM caspase inhibitor zVAD or 130 mM extracellular KCl. (**B**) THP1 cells were treated for 6 hours with 0.2 mg/ml asbestos, 0.5 mg/ml silica, or 0.2 mg/ml MSU. Particles and/or fibers entering cells are marked with arrows and fibers or particles on cell surface with arrowheads. (**C**) THP1 cells were stimulated for 6 hours with 0.2 mg/ml asbestos, 0.2 mg/ml MSU, 10 μg/ml R837, or 3.4 μM nigericin, in the absence or presence of 0.2 μM cytochalasin D. (**D**) Murine peritoneal macrophages were stimulated with asbestos, MSU, or ATP, in the presence of N-acetyl-L-cysteine (NAC) (25 mM) or APDC (50 μM). (**E**) THP1 cells were stimulated with asbestos (0.1 mg/ml) or MSU (0.1 mg/ml), in the presence or absence of deferoxamine mesylate (2 mM) or uricase (0.1 U/ml), respectively. (**F**) THP1 control cells (WT) or knock-down for the NADPH oxidase subunit p22phox were stimulated with 0.1 mg/ml asbestos, 0.1 mg/ml MSU or 0.01 mg/ml R837. (**G**) THP1 control cells (mock) or cells knocked down by shRNA against TRX were stimulated for 6 hours with 0.1 mg/ml asbestos, 0.1 mg/ml MSU, or 0.5 mg/ml silica. Media supernatants (SN) or cell extracts (Cell) were analyzed by Western blotting as indicated in the text.

Nalp3$^{-/-}$ asbestos-exposed mice (Fig. 4B). Several cytokines were induced by asbestos, such as IL-1β, KC, and IL-12(p40) (Fig. 4C). KC, a potent neutrophil chemoattractant, and IL-1β were decreased in Nalp3$^{-/-}$ mice, whereas IL-12(p40) production was independent of Nalp3. Although mucin production, which parallels inflammatory changes in lungs and may occur to facilitate clearance of asbestos fibers (21), was similarly increased by asbestos in Nalp3$^{-/-}$ and Nalp3$^{+/+}$ lungs, significant increases in the number of mucin-producing distal bronchioles, sites of impact of chrysotile asbestos fibers after inhalation, occurred in Nalp3$^{-/-}$ mice (table S1).

The Nalp3 inflammasome is implicated in the pathological increase of IL-1β production in

autoinflammatory syndromes, such as Muckle-Wells syndrome (22), as well as inflammatory processes, such as gout and pseudogout (7). Our findings support the implication of the Nalp3 inflammasome in pulmonary inflammatory diseases that are linked to pathogenic air pollutants and can ultimately lead to lung cancer and fibrosis. Asbestos, MSU, and probably other particles activate the NALP3 inflammasome in a similar way, requiring actin-mediated cellular uptake in contrast to small, nonparticulate molecules such as R837, ATP, or nigericin. How the phagocytosed fibers, particles, and MSU crystals are sensed by the Nalp3 inflammasome is not completely clear. It seems unlikely, however, that each of the particles is "specifically" recognized

by Nalp3. Rather, our data support a model in which ROS generated by a NADPH oxidase are implicated in Nalp3 inflammasome activation (Fig. 4D). ROS constitute one of the most ancient danger signals and are generated in large amounts by NADPH oxidase after microbe (and fine particle) phagocytosis.

An important role for IL-1β has been proposed in the pathogenesis of asbestos-induced mesothelioma because it regulates human mesothelial cell proliferation (23), and IL-1β-driven inflammation is well known to promote the development and invasiveness of several tumor types in vivo (24). Moreover, in a mouse model of bleomycin- or silica-induced pulmonary fibrosis, treatment with IL-1 receptor antagonist

Fig. 4. In vivo inhalation of asbestos results in reduced pulmonary inflammation in Nalp3-deficient mice. (A) Total and (B) differential cell counts in BALF in 9-day sham- and asbestos-exposed groups. (C) Cytokine and/or chemokine levels were measured in BALF of 9-day asbestos-exposed mice. Values are means ± SEM. *Significantly different from sham of same genotype. †Significantly different from wild type. (D) A proposed model for asbestos-induced inflammasome activation. See text for details.
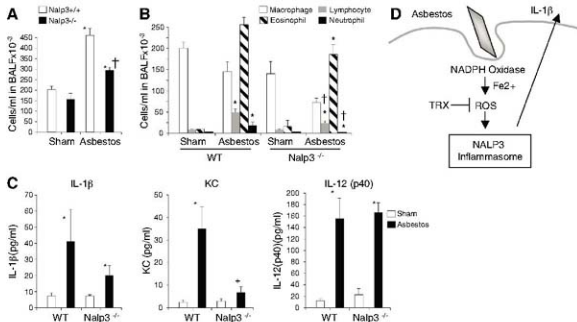
(IL-1ra) reduces the proportion of damaged lung (25). Silicosis and asbestosis continue to be a common cause of chronic lung disease, despite evidence that they can be prevented by environmental dust control.

Anakinra (IL-1ra) is efficiently used in the clinical treatment of autoinflammatory syndromes (26), as well as for gout patients (27). The present study suggests that Anakinra may be able to slow down progression of asbestosis, silicosis, and possibly other inflammatory lung diseases.

References and Notes

1. B. T. Mossman, A. Churg, Am. J. Respir. Crit. Care Med. 157, 1666 (1998).
2. M. E. Ramos-Nino et al., J. Cell. Biochem. 98, 723 (2006).
3. D. W. Kamp, S. A. Weitzman, Thorax 54, 638 (1999).
4. B. T. Mossman, J. Bignon, M. Corn, A. Seaton, J. B. Gee, Science 247, 294 (1990).
5. B. Rimal, A. K. Greenberg, W. N. Rom, Curr. Opin. Pulm. Med. 11, 169 (2005).
6. Materials and methods are available as supporting material on Science Online.
7. F. Martinon, A. Mayor, A. Tardivel, J. Tschopp, Nature 440, 237 (2006).
8. F. Martinon, K. Burns, J. Tschopp, Mol. Cell 10, 417 (2002).
9. S. Mariathasan et al., Nature 440, 228 (2006).
10. F. Martinon, L. Agostini, E. Meylan, J. Tschopp, Curr. Biol. 14, 1929 (2004).
11. E. A. Miao et al., Nat. Immunol. 7, 569 (2006).
12. A. Amer et al., J. Biol. Chem. 281, 35217 (2006).
13. V. Petrilli et al., Cell Death Differ. 14, 1583 (2007).
14. L. Franchi, T. D. Kannenganti, A. R. Dubyak, G. Nuñez, J. Biol. Chem. 282, 18810 (2007).
15. T. Fernandes-Alnemri et al., Cell Death Differ. 14, 1590 (2007).
16. K. Hansen, B. T. Mossman, Cancer Res. 47, 1681 (1987).
17. A. Shukla et al., Free Radic. Biol. Med. 34, 1117 (2003).
18. C. M. Cruz et al., J. Biol. Chem. 282, 2871 (2007).
19. A. Haegens et al., J. Immunol. 178, 1800 (2007).
20. R. F. Robledo et al., Am. J. Pathol. 156, 1307 (2000).
21. T. Sabo-Attwood et al., Am. J. Pathol. 167, 1243 (2005).
22. L. Agostini et al., Immunity 20, 319 (2004).
23. Y. Wang et al., Int. J. Oncol. 25, 173 (2004).
24. Y. Krelin et al., Cancer Res. 67, 1062 (2007).
25. P. F. Piguet, C. Vesin, G. E. Grau, R. C. Thompson, Cytokine 5, 57 (1993).
26. P. N. Hawkins, H. J. Lachmann, M. F. McDermott, N. Engl. J. Med. 348, 2583 (2003).
27. A. So, T. De Smedt, S. Revaz, J. Tschopp, Arthritis Res. Ther. 9, R28 (2007).
28. We thank T. Barrett and M. MacPherson for technical assistance with animal experiments; K. Butnor for pathology analysis; M. MacPherson, V. Alexeeva, and M. von Turkovich for performing SEM analyses; E. Logette, D. Muruve, and F. Martinon for helpful discussions; P. Vandenabeele, University of Ghent, for the antibody against caspase1; and V. Dixit (Genentech, San Francisco) for the ASC−/− and Ipaf−/− mice. This work was supported by a NIH Program Project grant from the National Heart Lung and Blood Institute (P01HL67004) to B.T.M. and a grant for work in molecular oncology from the Commission for Technology and Innovation, National Centers of Competence in Research (CTI, NCCR),

A Haptoglobin-Hemoglobin Receptor Conveys Innate Immunity to *Trypanosoma brucei* in Humans

Benoit Vanhollebeke,[1] Géraldine De Muylder,[1] Marianne J. Nielsen,[2] Annette Pays,[1] Patricia Tebabi,[1] Marc Dieu,[3] Martine Raes,[3] Soren K. Moestrup,[2] Etienne Pays[1]*

The protozoan parasite *Trypanosoma brucei* is lysed by apolipoprotein L-I, a component of human high-density lipoprotein (HDL) particles that are also characterized by the presence of haptoglobin-related protein. We report that this process is mediated by a parasite glycoprotein receptor, which binds the haptoglobin-hemoglobin complex with high affinity for the uptake and incorporation of heme into intracellular hemoproteins. In mice, this receptor was required for optimal parasite growth and the resistance of parasites to the oxidative burst by host macrophages. In humans, the trypanosome receptor also recognized the complex between hemoglobin and haptoglobin-related protein, which explains its ability to capture trypanolytic HDLs. Thus, in humans the presence of haptoglobin-related protein has diverted the function of the trypanosome haptoglobin-hemoglobin receptor to elicit innate host immunity against the parasite.

P arasites need to evade both the innate and acquired immunity of their hosts, and this process results from continuous evolution of mutual self-defense mechanisms. Such is the case of human resistance and sensitivity to diff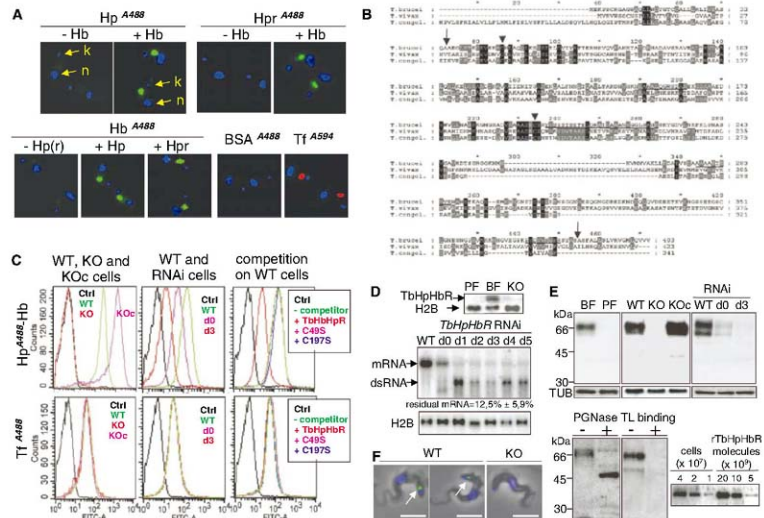erent subspecies of the African trypanosome *Trypanosoma brucei*. Apolipoprotein L-I (apoL1) is a primate-specific serum apolipoprotein that mainly exists bound to a subset of high-density lipoprotein (HDL) particles that also contain

haptoglobin-related protein (Hpr) (1–3). ApoL1 protects humans against infection by *T. brucei*, with the exception of subspecies that cause sleeping sickness (*T. brucei* subsp. *rhodesiense* and *T. brucei* subsp. *gambiense*) (4, 5). Trypanosome lysis results from the routing of apoL1 into the lysosome (4–7), where this protein triggers uncontrolled vacuole swelling due to its anion-selective

pore-forming activity in the lysosomal membrane (7, 8). ApoL1 uptake by the parasite is mainly mediated by the Hpr component of the carrier HDL particles termed trypanosome lytic factor 1 (TLF-1) (9, 10). In addition, a minor apoL1 fraction present in distinct Hpr-containing complexes termed TLF-2 enters trypanosomes independently of Hpr (11, 12). Hpr is a primate-specific protein sharing 91% sequence identity with haptoglobin (Hp) (13). Both proteins bind hemoglobin (Hb) with high affinity (14). However, although Hp-Hb complexes are specifically recognized by the monocyte/macrophage-specific scavenger receptor CD163 for efficient clearance of Hb from blood, Hpr-Hb complexes do not bind to this receptor (14, 15).

Hp has been shown to compete equally as well as Hpr for the uptake of TLF-1 by *T. brucei*, suggesting that the parasite receptor for these particles is unable to discriminate between Hpr and Hp (9, 10). Accordingly, in the present study we observed that both Hp and Hpr were taken up by the parasite, although this required Hb. Conversely, Hb was internalized only when present together with Hp or Hpr (Fig. 1A). To identify the *T. brucei* receptor for Hp-Hb/TLF-1, trypanosome extracts were submitted to affinity chromatography on resins containing immobilized Hp-Hb complexes (16). Analysis by mass spectrometry revealed the product of gene *Tb927.6.440*, a putative glycosyl-phosphatidyl inositol (GPI)–anchored

[1]Laboratory of Molecular Parasitology, Institute for Molecular Biology and Medicine, Université Libre de Bruxelles, 12 rue des Profs Jeener et Brachet, 86041 Gosselies, Belgium. [2]Department of Medical Biochemistry, University of Aarhus, DK-8000 Aarhus, Denmark. [3]Unit of Cellular Biochemistry and Biology, University of Namur, 85000 Namur, Belgium.

*To whom correspondence should be addressed. E-mail: epays@ulb.ac.be

Fig. 1. Identification of TbHpHbR. (**A**) Lysosomal accumulation of Alexa Fluor 488–conjugated Hp, Hpr, Hb, bovine serum albumin (BSA), and Alexa Fluor 594–conjugated Tf, incubated individually or with nonlabeled binding partners in the presence of the protease inhibitor FMK-024. BSA and Tf are respectively negative and positive controls for specific uptake. The blue dots represent 4',6'-diamidino-2-phenylindole–stained DNA (k, kinetoplast; n, nucleus). (**B**) Alignment of amino acid sequences of Tb927.6.440 and homologs in *T. vivax* and *T. congolense*. Arrows indicate predicted cleavages of N- and C-terminal signal peptides in *T. brucei*. Arrowheads point to conserved cysteines. The underlined peptides were identified by mass spectrometry. (**C**) Flow cytometry analysis of trypanosomes incubated w/o (Ctrl) with Alexa Fluor 488–conjugated Hp-Hb or Tf, in the presence or not of a fivefold excess of recombinant TbHpHbR or TbHpHbR mutants (C49S and C197S). RNAi was

with or without the addition of doxycyclin (Dox) for 3 days. (**D**) Detection of *TbHpHbR* transcripts from procyclic forms (PF), bloodstream forms (BF), and KO and RNAi cells. dsRNA, double-stranded RNA; d, days after Dox induction; histone H2B mRNA was the loading control. (**E**) Immunodetection of TbHpHbR in Hp-Hb binding extracts from various trypanosome lines. Tubulin (TUB) was the Hp-Hb column-loading control. Extracts from WT BFs were treated (+) or not (−) with peptide-N-glycosidase F (PNGase). TbHpHbR was also immunodetected in WT BF extracts bound to tomato lectin (TL) after chromatography with (+) or without (−) competing chito-oligosaccharides. (Bottom right) Comparative TbHpHbR immunodetection in deglycosylated extracts and reference recombinant protein amounts. (**F**) Localization of TbHpHbR by immunofluorescence of WT and KO BFs. Arrows highlight the fluorescent spots. Scale bar, ~10 μm.

surface protein that displays no homology with sequences of known function (Fig. 1B). In *T. brucei* subsp. *brucei*, the single gene Tb927.6.440 was located at the end of a polycistronic transcription unit on chromosome 6 (fig. S1). It was also present in *T. brucei* subsp. *gambiense* and *T. brucei* subsp. *rhodesiense* but absent from the related kinetoplastids *T. cruzi* and *Leishmania*. In *T. vivax* and *T. congolense*, putative GPI-anchored proteins showed significant similarity [expectation values $= 9.7 \times 10^{-10}$ and 4.9×10^{-16}, respectively] (Fig. 1B). Thus, Tb927.6.440-like sequences were present only in kinetoplastids that exhibited entire development in the bloodstream of their hosts. The involvement of Tb927.6.440 in the binding and uptake of the Hp-Hb complex was next examined using gene knockdown via RNA interference (RNAi), gene knockout (KO), and competition with excess recombinant protein (Fig. 1, C to F, and fig. S2). Under each of these conditions, the uptake of Hp-Hb was inhibited, whereas that of the control protein transferrin (Tf) was not affected (Fig. 1C and fig. S3). Moreover, reintroduction of Tb927.6.440 by transfection into KO cells [KO complemented (KOc)] restored the uptake of Hp-Hb (Fig. 1C and fig. S3). Thus, experimental evidence suggested

that Tb927.6.440 represented the *T. brucei* Hp-Hb receptor (TbHpHbR).

Expression analysis demonstrated that TbHpHbR is a bloodstream stage-specific protein present at 200 to 400 copies per cell (Fig. 1, D and E). The apparent molecular weight (72 kD) was higher than predicted (43.3 kD) because of N-glycosylation (Fig. 1E). The binding of TbHpHbR to tomato lectin (Fig. 1E) revealed the presence of linear chains of poly-*N*-acetyllactosamine, a hallmark of proteins from the endocytic pathway (*17*). As was observed for other *T. brucei* receptors (*18*), TbHpHbR localized close to the kinetoplast, in the flagellar pocket region (Fig. 1F).

We next examined the binding parameters of TbHpHbR using surface plasmon resonance (Fig. 2, A to C, and fig. S4). TbHpHbR demonstrated binding properties similar to those of the human functional analog CD163 because it recognized the Hp-Hb complex with high affinity [dissociation constant (K_d) $\sim 13 \times 10^{-9}$ M] but none of the proteins individually. Unlike CD163, however, TbHpHbR bound Hpr-Hb as well as Hp-Hb ($K_d \sim 17 \times 10^{-9}$ M) (Fig. 2A), and the two receptors responded differently to changes in pH and Ca²⁺ (Fig. 2B). It seemed likely that the

ligand-binding site of TbHpHbR was dependent on a disulfide bridge, because the replacement of either of two conserved cysteines (residues 49 and 197, Fig. 1B) by serine abolished the binding (Fig. 1C). The binding properties of CD163 and TbHpHbR were similar but not identical, because binding of Hp-Hb to CD163 was inhibited by excess TbHpHbR but not to the same extent as seen for the extracellular (soluble) domain of CD163 (Fig. 2C). Of the two TbHpHbR mutants significantly competed in these experiments (Fig. 2C). In vivo, Hp-Hb appeared to be the only essential ligand for TbHpHbR because in Hp⁻/⁻ mice the growth rates of TbHpHbR KO and wild-type (WT) trypanosomes were similar, whereas in Hp⁺/⁺ animals, KO parasites grew significantly slower than WTs (Fig. 2D).

Recombinant TbHpHbR, but not the C49S (*19*) mutant, was able to bind TLF-1 in affinity binding assays (*16*) (Fig. 3A). Furthermore, loss of TbHpHbR after RNAi or gene KO (Fig. 1, D to F, and fig. S2) conferred resistance to human HDL-mediated lysis. The lytic activity of normal human serum (NHS) on these parasites was decreased about 200-fold and was comparable to that observed on WT parasites when NHS saturated with competing Hp was used (Fig. 3B). Accordingly, KO parasites were refractory to lysis by TLF-1, but not to lysis by TLF-2, for which Hpr is not involved in uptake (Fig. 3C). In contrast, overexpression of TbHpHbR in KOc cells (Fig. 1, C and E) led to increased sensitivity to NHS (Fig. 3B). Finally, the role of TbHpHbR in TLF-1–mediated lysis was also confirmed by competition experiments in which TLF-1 activity was blocked equally well by excess recombinant TbHpHbR or Hb complexes with recombinant Hp or Hpr but not by Hp or Hpr alone (Fig. 3D).

Trypanosomes seemed to require Hp-Hb because growth rates were reduced in Hp⁻/⁻ mice (Fig. 2D) or in anhaptoglobinemic [Hp(r)⁻/⁻] human serum (*10*, *20*), but were restored to normal by addition of exogenous Hp-Hb (Fig. 4A). TbHpHbR RNAi or KO cells also grew slower than WT or KOc cells (Figs. 2D and 4A). The growth-promoting effect of Hp-Hb was probably not linked to Hp-derived iron because in *T. brucei* iron is internalized through a specific iron-dependent Tf receptor (*18*), and Tf uptake was not up-regulated in the absence of the Hp-Hb receptor (Fig. 1C). Although African trypanosomes are deficient in heme biosynthesis (*21*), *T. brucei* bloodstream forms contain hemoproteins, such as cytochromes P450 and b5 (*22–26*). Therefore, Hb uptake could satisfy the heme requirement of the parasite. Uptake studies of Hp-Hb, either ¹²⁵I-labeled or containing ¹⁴C-labeled heme, performed in WT, KO, and TbHpHbR overexpressor KOc trypanosomes in the presence or absence of the lysosomal protease inhibitor FMK-024, indicated that TbHpHbR allows the intracellular accumulation of heme, whereas the protein carrier undergoes fast degradation (Fig. 4B). Heme accumulation saturated at around 6.5 ng/mg of protein and remained durably cell-associated,
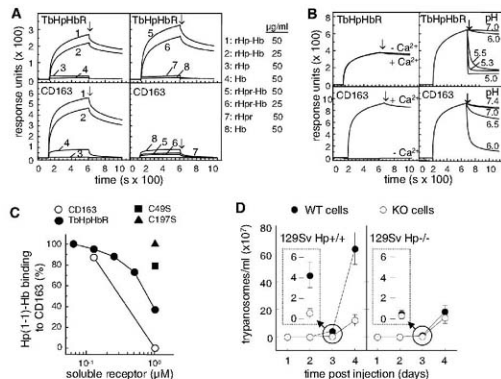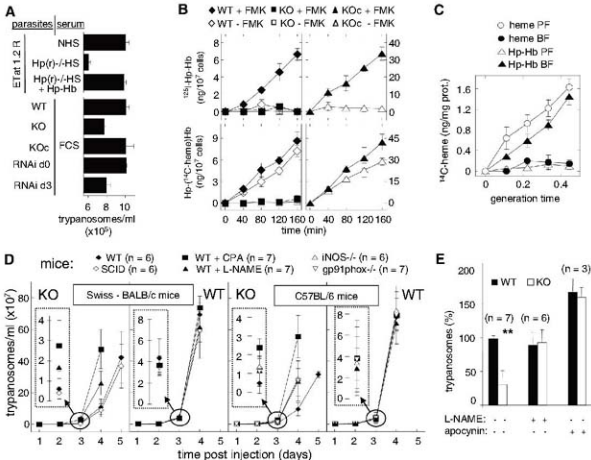

Fig. 2. Binding properties of TbHpHbR. (**A** to **C**) Surface plasmon resonance analysis. (A) Complexes of purified human Hb with purified recombinant Hp (rHp, phenotype 1-1) or Hpr were analyzed for binding to purified recombinant TbHpHbR immobilized on a Biacore sensor chip. For comparison, complexes of Hb with rHp or rHpr were analyzed for binding to immobilized CD163 purified from human spleen. Similar results were obtained with purified Hp (either 1-1 or 2-2 phenotype). (B) Binding of Hp-Hb (25 μg/ml) to immobilized TbHpHbR or CD163, in the presence or absence of 2 mM Ca²⁺ in the flow buffer, or at various pH values. The association phase was recorded at pH 7.4. The arrows indicate the time points for the recording of the dissociation phase, and the pH change. (C) Binding of Hp-Hb (25 μg/ml) to immobilized human CD163 in the presence of increasing concentrations of soluble CD163, TbHpHbR, or TbHpHbR mutants. The effect of soluble receptor on the binding of the Hp-Hb to immobilized CD163 is shown relative to the binding recorded in the absence of soluble receptor. rHp and purified Hp were from the 1-1 haplotype. (**D**) Parasitemia by WT and TbHpHbR KO trypanosomes in Hp⁺/⁺ and Hp⁻/⁻ 129Sv mice.
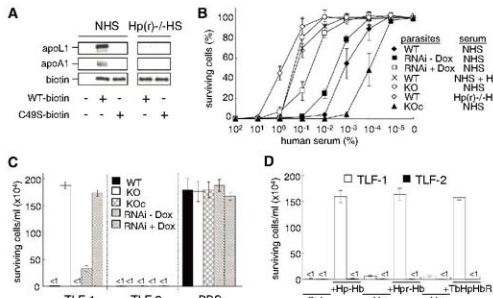
Fig. 3. Involvement of TbHpHbR in trypanolysis. (**A**) Evidence that TbHpHbR binds TLF-1. NHS and anhaptoglobinemic human serum [Hp(r)$^{-/-}$HS] were eluted through streptavidin-agarose beads containing, or not, biotin-conjugated recombinant WT TbHpHbR or TbHpHbR mutant. The bound fraction was analyzed for the presence of apoA1 and apoA1 (TLF-1 markers). (**B**) Trypanolysis assays with human serum. Trypanosomes (10^5/ml) were incubated in HMI-9 medium containing the indicated concentrations of serum. Surviving cells were counted after 24 hours. Dox, 1 μg/ml; Hp, 200 μg/ml. (**C**) Trypanolysis assays with isolated TLF-1 and TLF-2 complexes. The fractions were incubated with 10^6 trypanosomes/ml. Surviving cells were counted after 6 hours. (**D**) Same as in (C), but on WT cells in the presence or absence of the indicated proteins (200 μg/ml each).

Fig. 4. In vivo function of TbHpHbR (unless indicated, error bars indicate standard deviation from four independent experiments). (**A**) Cell density of various trypanosome lines after 24 hours of incubation in vitro in sera depleted or not of Hp. Etat 1.2R, *T. brucei* subsp. *rhodesiense* Edinburgh Trypanosoma antigen type 1.2R; d, days; FCS, fetal calf serum. (**B**) Accumulation of radiolabeled Hp-Hb [either ^{125}I–(Hp-Hb) or Hp–(^{14}C-heme)Hb] in 10^7 trypanosomes, monitored in the presence or absence of the lysosomal protease inhibitor FMK-024. (**C**) Accumulation of ^{14}C-heme (ng/mg of protein) from either Hp–(^{14}C-heme)Hb or free ^{14}C-heme in BFs and PFs. The cell doubling time was 6 hours and 18 hours for BFs and PFs, respectively. (**D**) Parasitemia of mice injected intraperitoneally with 10^4 WT or KO trypanosomes, with or without prior administration of cyclophosphamide (CPA) or L-NAME. SCID mice (BALB/c background) were injected as controls. The absence of data at day 5 means that all mice died between days 4 and 5. (**E**) Effect of peritoneal exudate cells (PECs) from Swiss mice on growth of WT and KO parasites, after 48 hours of incubation in the presence or absence of L-NAME (500 μM) or apocynin (250 μM). Under the conditions used, L-NAME reduced NO synthesis by 58.3 ± 16% ($n = 5$ independent experiments). The trypanosome density was normalized to that of WT cells incubated with nonstimulated PECs. Asterisks indicate significant differences based on unpaired Student's *t* test (**, $P < 0.01$).

with a loss of 16% per generation time (fig. S5A). The steady-state heme content of WT cells isolated from mice was 2.3 ng/mg of protein, whereas heme was undetectable in KO cells (16). In trypanosome lysates, heme appeared to be mostly incorporated into hemoproteins because it was recovered in acetone-insoluble material, as occurs with hemoproteins such as Hb but not with free heme (fig. S5B). Accordingly, after uptake at least two ^{14}C-labeled bands could be detected after electrophoresis of protein extracts (fig. S5C). Free ^{14}C-labeled heme was not internalized in bloodstream forms, even when provided with TbHpHbR, whereas it accumulated in the insect-specific procyclic forms that lack TbHpHbR and do not internalize Hp-Hb (Figs. 1D and 4C). Altogether, these data indicate that TbHpHbR directs the internalization of heme carried by the Hp-Hb complex into hemoproteins in order to sustain growth of bloodstream forms. In mice (Swiss or C57BL/6), the growth reduction of KO parasites was alleviated after pretreatment of the animals with the general immunosuppressor drug cyclophosphamide (Fig. 4D). However, this inhibition was conserved in lymphocyte-deprived, severe combined immunodeficient (SCID) mice (BALB/c background) (Fig. 4D). Therefore, we investigated the sensitivity of trypanosomes to macrophages. When incubated with peritoneal macrophages, KO

parasites were more affected than WTs (Fig. 4E). This inhibitory effect was substantially relieved when inhibitors of the key enzymes of macrophage oxidative burst were added, either the NO synthase (a nitric oxide producer) inhibitor L-N^G-nitroarginine methyl ester (L-NAME) or the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (a superoxide anion producer) inhibitor apocynin [1-(4-hydroxy-3-methoxyphenyl)ethanone] (Fig. 4E). None of these drugs influenced trypanosome growth in vitro (fig. S6). In support of these data, parasitaemia by TbHpHbR KO parasites was improved to a similar extent either after injection of L-NAME in WT mice or when inducible nitric oxide synthase (iNOS)$^{-/-}$ mice or NAPDH oxidase (gp91phox)$^{-/-}$ mice were used (Fig. 4D). Therefore, TbHpHbR appeared to confer increased resistance to oxidative stress induced by macrophages.

These data suggest that African trypanosomes have evolved a receptor specifically designed to acquire heme from Hp-Hb for incorporation into hemoproteins that both increase the trypanosome's growth rate and resistance to the oxidative response of the host. The mechanism of the resistance to oxidative attack is yet unknown but could involve hemoprotein-mediated modification of membrane lipids (23, 25). In contrast to human CD163, TbHpHbR recognizes Hp-Hb and Hpr-Hb complexes equally well. Therefore,

the presence of Hpr on human lytic HDL particles triggered internalization of the Hb-exposed fraction of these particles, which also contain the trypanolytic factor apoL1. This finding now explains the stimulating effect of Hb on trypanolysis as an indirect mechanism (27). Thus, in human serum the presence of the Hp-Hb receptor became detrimental instead of protective. In turn, synthesis of the apoL1 physical inhibitor serum resistance–associated protein allowed *T. brucei* subsp. *rhodesiense* to escape this problem and cause human sleeping sickness (4, 28).

References and Notes

1. P. N. Duchateau *et al.*, *J. Biol. Chem.* **272**, 25576 (1997).
2. A. M. Shiflett *et al.*, *J. Biol. Chem.* **280**, 32578 (2005).
3. E. B. Lugli *et al.*, *Mol. Biochem. Parasitol.* **138**, 9 (2004).
4. L. Vanhamme *et al.*, *Nature* **422**, 83 (2003).
5. E. Pays *et al.*, *Nat. Rev. Microbiol.* **4**, 477 (2006).
6. K. M. Hager *et al.*, *J. Cell Biol.* **126**, 155 (1994).
7. D. Pérez-Morga *et al.*, *Science* **309**, 469 (2005).
8. B. Vanhollebeke *et al.*, *J. Eukaryot. Microbiol.* **54**, 448 (2007).
9. J. Drain *et al.*, *J. Biol. Chem.* **276**, 30254 (2001).
10. B. Vanhollebeke *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4118 (2007).
11. J. Raper *et al.*, *J. Exp. Med.* **183**, 1023 (1996).
12. J. Raper *et al.*, *Infect. Immun.* **67**, 1910 (1999).
13. N. Maeda, *J. Biol. Chem.* **260**, 6698 (1985).
14. M. J. Nielsen *et al.*, *Blood* **108**, 2846 (2006).
15. M. Kristiansen *et al.*, *Nature* **409**, 198 (2001).
16. Materials and methods are available as supporting material on *Science* Online.
17. D. P. Nolan *et al.*, *Curr. Biol.* **9**, 1169 (1999).
18. D. Steverding, *Microbes Infect.* **8**, 2777 (2006).
19. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
20. Y. Koda *et al.*, *Am. J. Hum. Genet.* **62**, 245 (1998).
21. M. Berriman *et al.*, *Science* **309**, 416 (2005).
22. B. J. Berger, A. H. Fairlamb, *Biochem. Pharmacol.* **46**, 149 (1993).
23. K. E. Tripodi *et al.*, *FEBS J.* **273**, 271 (2006).
24. R. Broadhead *et al.*, *Nature* **440**, 224 (2006).
25. G. I. Lepesheva *et al.*, *Chem. Biol.* **14**, 1283 (2007).
26. D. J. Bridges *et al.*, *Proteomics* **8**, 83 (2008).
27. J. Widener *et al.*, *PLoS Pathog.* **3**, 1250 (2007).
28. H. V. Xong *et al.*, *Cell* **95**, 839 (1998).
29. We thank C. Jacobsen, L. Hederstedt, M. Boutry, A. Beschin, A. Jacquet, S. Denanglaire, A. M. Bundsgaard, E. Dupont, and D. Nolan for advice and help, and P. Brouckaert and E. Tolosano for iNOS$^{-/-}$ and Hp$^{-/-}$ mice, respectively. This work was supported by the Belgian National Fund for Scientific Research, the J. Brachet Fund, the Danish Medical Research Council, and the Interuniversity Attraction Poles Programme—Belgian Science Policy. B.V. and G.D. are research fellows at the Fonds National de la Recherche Scientifique and Fonds de la Recherche pour l'Industrie et l'Agriculture, respectively.

Sample Work-Up Station

The Carousel Work-Up Station facilitates parallel or sequential work-up of up to 12 samples, using filtration, phase separation, liquid/liquid extraction, or solid-phase extraction techniques. It is designed for use with the 12-position Carousel Reaction Station or similar parallel synthesizer. The station accepts up to 12 70-ml columns loaded into one of two identical stackable racks. It enables collection of product into the lower rack with Carousel Reaction Tubes or standard 1-inch boiling tubes. The unique design requires no taps, valves, or drip needles, making assembly and operation fast and simple. The system can be used under gravity or with the innovative SpeedFlow Booster, allowing precise individual control of solvent flow within each column.

Radleys
For information +44 1799-513320
www.radleys.com

Services to Profile microRNAs

A collection of new services to profile microRNAs (miRNAs) includes a new high-sensitivity assay for Asuragen's DiscovArray miRNA Expression Profiling Service, and expansion of Asuragen's miRInform miRNA data packaging system to support the Agilent miRNA array data, the new ABI Taqman(a) Human v 1.0 miRNA Array, and a new miRNA-compatible sample prep service for PAXgene blood samples. The DiscovArray Service makes use of an Affymetrix-based chip that contains more than 13,000 candidate and Sanger miRNAs and requires only 50–200 ng total RNA input. It can profile even from challenging samples such as formalin-fixed paraffin embedded tissues. miRInform is for the normalization and delivery of miRNA data in DVD format with interactive figures that link directly to Sanger mirBase, tables, and publication-ready images. miRInform is a cost-effective tool for mining data from the Agilent Human miRNA Microarray without requiring third-party tools or extensive statistical training, and comes with a personal data review with a trained analyst.

Asuragen
For information 512-681-5200
www.asuragen.com

Oligonucleotide Database

The Oligome is a comprehensive database of more than 10 million oligonucleotide probes designed to the latest release of the human genome. The Oligome makes it possible to provide custom-designed oligonucleotide arrays for array-based comparative genomic hybridization on request in a short time. Oxford Gene Technology's 60MER oligonucleotide microarrays are fabricated using inkjet technology with base-by-base synthesis that allows high-precision feature placement and denser coverage of regions of interest. The company offers custom-designed comparative genomic hybridization arrays for analyzing chromosomal abnormalities within focused areas of the human genome at high resolution. As part of the free design service, scientists can simply provide their regions of interest, and the company then performs the Oligome-based bioinformatics required to design, fabricate, and deliver the desired focused arrays.

Oxford Gene Technology
For information +44 1865-856828
www.ogt.co.uk

Large Scale Imaging Confocal Microscope

The Leica TCS LSI (large scale imaging), the first macro-zoom confocal microscope, offers a unique combination of high-resolution imaging plus a large field "macro" view for in vivo imaging of model organisms. The Leica TCS LSI provides researchers with the tool they need to visualize and precisely image the development of life and the complex cellular interactions within whole, living animals. As organisms grow, these studies require an imaging system that provides high resolution, a large workspace, and a wide field of view to reveal the dynamics of cell growth, cell differentiation processes, and the development of organs in vivo. The system's 16:1 optical zoom offers both the advantage of high magnification and high resolution for true spectral confocal imaging and the ability to seamlessly move to a macro view of the entire specimen without changing any hardware. Such a system is suitable for revealing the finest details in *Drosophila*, *Arabidopsis*, mouse, or zebrafish. The optical zoom system is fully apochromatic, the highest correction available for image quality and color rendition. The instrument makes it easier for scientists to identify new pathways from gene to cell or from cell to animal, analyze protein interactions, test the influence of drugs in vivo, and examine the influence of genomic defects of the whole animal.

Leica
For information 800-248-0123
www.leica-microsystems.com

Human Natural Killer Cells

The Poietics Human Natural Killer Cells provide research laboratories with cryopreserved, ready-to-use primary cells for immunology research in areas such as immune cell activation and regulation, cancer, autoimmune disease/human immunodeficiency virus, viral infection, vaccine development, immunotherapy, and transplantation. The cells are isolated from normal human peripheral blood using positive or negative immunomagnetic selection for the CD56 antigen. These cells are tested to be at least 90% pure for CD56+ via flow cytometry, and at least 95% viable after being thawed. This highly homogeneous population of natural killer cells yields cleaner, more consistent results. Matched sets of other immune cells from the same donor are available as well.

Lonza Group
For information 800-638-8174
www.lonza.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.